UNIVERSITÉ DE FRIBOURG
UNIVERSITÄT FREIBURG

WORKING PAPERS SES

Evaluating local average and quantile treatment effects under endogeneity based on instruments: a review

Martin Huber
and
Kaspar Wüthrich

N. 479
II.2017

Faculté des sciences économiques et sociales
Wirtschafts- und sozialwissenschaftliche Fakultät

Evaluating local average and quantile treatment effects under endogeneity based on instruments: a review

Martin Huber* Kaspar Wüthrich†

February 24, 2017

Abstract

This paper provides a review of methodological advancements in the evaluation of heterogeneous treatment effect models based on instrumental variable (IV) methods. We focus on models that achieve identification through a monotonicity assumption on the selection equation and analyze local average and quantile treatment effects for the subpopulation of compliers. We start with a comprehensive discussion of the binary treatment and binary instrument case which is relevant for instance in randomized experiments with imperfect compliance. We then review extensions to identification and estimation with covariates, multi-valued and multiple treatments and instruments, outcome attrition and measurement error, and the identification of direct and indirect treatment effects, among others. We also discuss testable implications and possible relaxations of the IV assumptions, approaches to extrapolate from local to global treatment effects, and the relationship to other IV approaches.

JEL Classification: C26

Keywords: instrument, LATE, treatment effects, selection on unobservables.

We have benefitted from comments by Anna Solovyeva and Svitlana Tyahlo.

*Corresponding author, University of Fribourg, Bd. de Pérolles 90, 1700 Fribourg, Switzerland, martin.huber@unifr.ch

†UC San Diego, Department of Economics, San Diego, 9500 Gilman Dr. La Jolla, CA 92093, USA, kwuthrich@ucsd.edu

1 Introduction

In empirical research, the assessment of the causal effect of a treatment (e.g. training or education) on an outcome of interest (e.g. earnings) is frequently complicated by endogeneity, implying that the treatment is not as good as randomly assigned. In other words, individuals may select themselves into the treatment in a non-random way that is related to their expected gains from the treatment in the outcome. This happens for instance in randomized experiments with non-compliance in which access to the treatment is randomly assigned, but some individuals decide not to comply with the randomization but choose a different treatment state. If compliance behaviour is associated with unobserved characteristics (e.g. motivation or ability) that also affect the outcome, endogeneity jeopardizes a causal analysis based on simple comparisons between treated and non-treated observations. In the presence of an instrumental variable (IV) that (i) affects the treatment decision of (at least) some subpopulation and (ii) is otherwise not associated with the potential outcomes under either treatment state, causal effects can nevertheless be identified. For this reason, IV methods have become a cornerstone of causal inference.

This paper reviews the methodological advancements in the IV-based evaluation of treatment effects. We focus on methods that allow treatment effects to be heterogenous, implying that the effectiveness of a treatment may vary across study subjects as a function of their observed and unobserved characteristics. In such models with a binary treatment and binary instrument and under the restriction that the treatment is weakly monotonic in the instrument, two stage least squares (TSLS) consistently estimates the average treatment effect for the compliant subpopulation. This effect is usually referred to as local average treatment effect (LATE). In the experimental context, compliers are those individuals whose treatment status is induced by the assignment. That is, they take-up the treatment when randomized in, but abstain from it when randomized out. Following the seminal paper of Imbens and Angrist (1994), much progress has been made in extending the initial framework in various empirically relevant dimensions. This includes for instance the evaluation of distributional and quantile treatment effects, multivalued or multiple treatments and instruments, identification and estimation in the presence of observed covariates, attrition and measurement error, and more. Furthermore,

it has been acknowledged that the LATE assumptions have testable implications that may be verified in the data and that specific causal effects might be point or partially identified under weaker conditions. Finally, conditions and tests for the external validity of the LATE with respect to the average treatment effect (ATE) in the total population have been proposed, which appears important in the light on the controversial debate in the literature about the empirical relevance of the complier population; see for instance the discussions in Deaton (2010), Imbens (2010a), Heckman and Urzúa (2010).

Our survey complements more introductory surveys of the LATE framework, see Imbens (2014) and the textbook discussions in Angrist and Pischke (2009) and Angrist and Pischke (2015). A more specialized review focussing on the specific aspects of identifying and estimating the local quantile treatment effect (LQTE) is provided by Melly and Wüthrich (2016).

We structure the review as follows. Section 2 reviews the IV assumptions in the binary instrument and treatment case and the identification of the LATE, LQTE, and potential outcome means and distributions. It also discusses identification under multivalued treatments and instruments and considers the concept of marginal treatment effects. Section 3 discusses a conditional version of the IV assumptions in the presence of covariates along with the identification of local, quantile, and marginal treatment effects as well as more general functionals among compliers. Section 4 discusses extensions of the IV framework to more complex identification problems, including non-response bias in the outcome, measurement errors in the treatment or the instrument, the presence of dynamic, i.e. sequentially assigned, or multiple treatments, and the evaluation of causal mechanisms (or direct and indirect effects) of the treatment. Section 5 discusses how violations of the IV assumptions affect identification and under which relaxations of the assumptions causal effects on specific subpopulations can nevertheless be obtained. Section 6 outlines approaches to test the IV assumptions and briefly discusses sensitivity checks and bounds analysis under specific violations of the assumptions. Section 7 is concerned with the external validity of the LATE for the entire popluation. It discusses potential checks for external validity based on observables, conditions for extrapolating the LATE to the ATE and along with testable implications, and partial identification of the ATE based on the IV assumptions and possibly further restrictions. Section 8 clarifies the relationship of the framework considered in this paper and other IV approaches suggested in the literature. Specifically, we

discuss the connection to the classical linear IV model with covariates and to the instrumental variable quantile regression model (Chernozhukov and Hansen, 2005). Section 9 concludes.

2 Identification and estimation without covariates

We first consider a setup with a binary treatment and a binary instrument. Section 2.1 discusses the IV assumptions, while Section 2.2 shows the identification of the LATE, LQTE, and the potential outcomes among compliers. Section 2.4 extends the initial framework to the case of a multivalued treatment, while Section 2.3 is concerned with multivalued instruments and introduces the concept of marginal treatment effect.

2.1 Assumptions

The leading case in the program evaluation literature is the assessment of the effect of some binary intervention or treatment D (with $D \in \{1, 0\}$). Examples include receiving ($D = 1$) or not receiving ($D = 0$) a labor market intervention like a job training, an educational intervention like private schooling, or a health intervention like a medical treatment. Y denotes the outcome on which the effect ought to be estimated, for instance, labor market success such as employment or earnings, which is measured at some point in time after the treatment. Under endogeneity, unobserved factors affect both D and Y such that treatment effects cannot be identified from simple comparisons of the treatment and the control group. However, if there exists an instrumental variable Z which is relevant in the sense that it influences the treatment status and valid in the sense that it is not associated with the unobserved factors and does not directly affect the outcome, treatment effects can be identified.

Our formal discussion is developed within the potential outcome framework (see for instance Rubin, 1974). Denote by $D(z)$ the potential treatment state that would occur if the instrument Z was exogenously set to some value z, and by $Y(d)$ the potential outcome for setting the treatment to some $d \in \{1, 0\}$. We will henceforth assume a binary instrument ($Z \in \{1, 0\}$), which for the time being simplifies the exposition, while Section 2.3 extends the framework to multi-valued Z. As an illustrative example, consider the experimental evaluation of a job training program in which Z and D denote the randomized assignment of and the actual participation status in

the training, respectively. In this context, $D(1)$ and $D(0)$ denote the potential participation states when randomized into or out of the job training. Similarly, $Y(1)$ and $Y(0)$ denote the potential outcomes (e.g. employment states) when participating and not participating in the training. For each subject, only one of the two potential outcomes and treatment states are observed, because the observed variables are defined as $Y = D \cdot Y(1) + (1 - D) \cdot Y(0)$ and $D = Z \cdot D(1) + (1 - Z) \cdot D(0)$. Consequently, causal effects cannot be identified without further assumptions.

Table 1: Definition of types

Types (T)	D(1)	D(0)	Notion
a	1	1	Always takers
c	1	0	Compliers
d	0	1	Defiers
n	0	0	Never takers

Even without any assumptions, the population can, however, be split into four treatment compliance types (denoted by $T \in \{a, c, d, n\}$) defined by the joint potential treatment states under $z = 1$ and $z = 0$, see the discussion in Angrist et al. (1996). As shown in Table 1, the compliers ($c : D(1) = 1, D(0) = 0$) react on the randomization as intended by the researcher and participate in the training when $z = 1$, while abstaining from it when $z = 0$. For the remaining three types, $D(z) \neq z$ for either $z = 1$, or $z = 0$, or both: The always takers ($a : D(1) = 1, D(0) = 1$) always take the training irrespectively of the instrument status, the never takers ($n : D(1) = 0, D(0) = 0$) are never treated, and the defiers ($d : D(0) = 1, D(0) = 1$) react counter-intuitively to randomization by participating in the treatment when randomized out, but not participating when randomized in. As either $D(1)$ or $D(0)$ remains unknown in the data, one cannot infer on any subject's type, which is a function of both potential treatment states. This implies that any subject with a particular observed combination of the treatment and the instrument may belong to one of two types, as summarized in Table 2.

Table 2: Observed subgroups and types

Observed values of Z and D	Potential types T
$\{Z = 1, D = 1\}$	belongs either to a or to c
$\{Z = 1, D = 0\}$	belongs either to d or to n
$\{Z = 0, D = 1\}$	belongs either to a or to d
$\{Z = 0, D = 0\}$	belongs either to c or to n

Therefore, comparing $E[Y|D = 1] - E[Y|D = 0]$ or $E[Y|D = 1, Z = z] - E[Y|D = 0, Z = z]$

(for $z \in \{1, 0\}$) does generally not yield any causal effect, as the mixture of types differs across D or (D, Z), respectively. The reason is that types generally have different distributions of unobservables which may confound the treatment and outcome. To convey the intuition, we consider the following nonparametric IV model:

$$Y = \phi(D, U), \quad D = \eta(Z, V), \tag{2.1}$$

ϕ and η denote general functions, while U and V are the unobserved terms (possibly scalars or vectors) which may be arbitrarily associated with each other, thus causing the treatment to be endogenous. Potential outcomes and treatment states are readily obtained by exogenously setting the treatment and the instrument to particular values d and z:

$$Y(1) = \phi(1, U), \quad Y(0) = \phi(0, U), \quad D(1) = \eta(1, V), \quad D(0) = \eta(0, V).$$

As $D(1) = \eta(1, V)$ and $D(0) = \eta(0, V)$ differ across types, i.e., they have *different* potential treatments for *same* values of z, the distribution of V must necessarily differ across types (as D is a function of Z and V only). Therefore, U also differs across types if it is associated with V. This can be easily illustrated by means of the following parametric model, which is a special case of the general IV model (2.1):

$$Y = \alpha + \beta D + U, \quad D = I\{\gamma + \delta Z \geq V\}. \tag{2.2}$$

α, γ are constants, β and δ slope coefficients, and $I\{\cdot\}$ the indicator function which is equal to one if its argument is satisfied and zero otherwise. Furthermore, U, V are assumed to be scalars for the sake of simplicity. For the compliers, it holds that $D(1) = I\{\gamma + \delta \geq V\} = 1, D(0) = I\{\gamma \geq V\} = 0$, so that the distribution of V satisfies $\gamma + \delta \geq V > \gamma$. Among always takers, however, $D(1) = I\{\gamma + \delta \geq V\} = 1, D(0) = I\{\gamma \geq V\} = 1$, so that $V \leq \gamma$. Consequently, unless U and V are independent, the treatment and the outcome are confounded. Treatment effects can therefore only be identified under additional assumptions on Z are satisfied, as outlined

below.

Note that the parametric model in (2.2) postulates homogeneous treatment effects due to the additive separability of D and U. However, there is generally no reason to believe that treatment effects are constant across individuals. One therefore typically prefers models that allow for heterogeneous treatment effects. That is, the impact of D on Y may vary across the values of other (unobserved) factors. Imbens and Angrist (1994) postulate the identifying assumptions for nonparametric IV models like (2.1), with the caveat that under effect heterogeneity, effects can generally only be obtained for the subpopulation of compliers. The assumptions impose (i) statistical independence between Z and the joint distribution of the potential treatment states and outcomes and (ii) weak monotonicity of the treatment in the instrument. Formally, the first assumption can be stated as follows:

Assumption 2.1 (Joint independence)**.** $Z \perp (D(1), D(0), Y(1), Y(0))$

The symbol "$\perp$" denotes independence. Assumption 2.1 implies two subconditions. First, the instrument must be random so that it is unrelated with factors affecting the treatment and/or outcome, implying that $(U, V) \perp Z$ holds in model (2.1). Therefore, not only the potential outcomes/treatment states, but also the types, which are defined by the joint potential treatment states, are independent of the instrument. Second, Z must not have a direct effect on Y other than through D, i.e., satisfy an exclusion restriction, which can be seen from the fact that the potential outcomes are only defined in terms of d rather than z and d. This holds by the model definitions in (2.1) and (2.2), because Z does not enter the equation of Y as explanatory variable. To make these two aspects explicit, Assumption 2.1 may be split into two conditions, see Angrist et al. (1996): (i) $Z \perp (D(1), D(0), Y(1,1), Y(1,0), Y(0,1), Y(0,0))$ and (ii) $Y(1,d) = Y(0,d) = Y(d)$ for $d \in \{1,0\}$ (exclusion restriction), where $Y(z,d)$ denotes a potential outcome defined in terms of both the instrument z and the treatment d.

Concerning the plausibility of Assumption 2.1 in empirical applications, note that in a successfully conducted experiment, the randomness of Z holds by construction. Furthermore, the exclusion restriction holds if mere assignment for instance to a training program does not have a direct effect on the outcome, e.g. through increased motivation or frustration due to being (not) offered the training. While Assumption 2.1 is plausible for instance in a medical trial

where individuals in the control group receive placebo treatments, it might be more dubious in so-called quasi-experimental settings. Taking the estimation of the returns to education (D) as an example, Angrist and Krueger (1991) suggest using quarter of birth as instrument (Z), as it is related to years of education through regulations concerning the school starting age, but arguably neither is driven by factors also affecting income nor has a direct effect on income. However, Bound et al. (1995) contest Assumption 2.1 in the context of quarter of birth instruments and present evidence that seasonal patterns of births are related to family income, physical and mental health, and school attendance rates, all of which may affect income. Furthermore, Buckles and Hungerman (2013) document large differences in maternal characteristics for births throughout the year (with winter births being more often realized by teenagers and unmarried women) based on U.S. birth certificate data and census data. A careful assessment of instruments that may appear plausible at the first glance is therefore in order, in particular when they are not randomly assigned by the researcher and no placebo treatments are given to the control group.

It is worth noting that when aiming to identify a mean effect like the LATE (see (2.15) below), full independence between Z and $Y(1), Y(0)$ as postulated in Assumption 2.1 can be replaced by the weaker mean independence restriction $E(Y(d)|T = t, Z = 1) = E(Y(d)|T = t, Z = 0) = E(Y(d)|T = t)$ for $d \in \{0, 1\}$ and $t \in \{a, c, d, n\}$. However, when distributional features like quantile treatment effects are of interest, (full) independence is required. From a practical perspective, the distinction between mean and full independence is often less relevant, as it is generally hard to think of scenarios in which mean independence holds, but the stronger full independence does not. For instance, if one is willing to assume that an instrument is mean independent of the potential hourly wage, it seems reasonable to also assume that it is mean independent of the log of potential hourly wage. As the latter variable is a (one-to-one) nonlinear transformation of the original potential outcome, this implies independence also with respect to higher moments. Therefore, strengthening mean to full independence often comes with little costs in terms of credibility such that we do not consider mean independence in the remainder of this paper.

Assumption 2.2 (Monotonicity)**.** $\Pr(D(1) \geq D(0)) = 1$ *or* $\Pr(D(1) \leq D(0)) = 1$

Assumption 2.2 says that the potential treatment state of any individual does either not decrease (positive monotonicity, $\Pr(D(1) \geq D(0)) = 1$) or not increase (negative monotonicity, $\Pr(D(1) \leq D(0)) = 1$) in the instrument. We will henceforth only consider the case of positive monotonicity, because the case of negative monotonicity is symmetric. Assumption 2.2 rules out the existence of defiers (type $T = d$), because for the latter group, $D(1) < D(0)$. As a consequence, always takers, never takes and compliers exhaustively partition the whole population. Note that this condition is implicit in parametric models like (2.2), where δ is a constant so that the effect of Z is homogeneous and V is a scalar unobservable. Again, this may not be the case in more general models.

Assumption 2.2 is satisfied by construction in randomized experiments with so-called one-sided non-compliance (see Bloom (1984)) and a first stage: if no subject randomized out of a job training can manage to "sneak into" the training anyway, then $\Pr(D(0) = 1) = 0$ such that defiers as well always takers do not exist. Even in many field experiments where $\Pr(D(0) = 1) > 0$, the presence of defiers appears implausible as it would imply counter-intuitive behavior to the randomization protocol. In several quasi-experimental settings, however, the assumption might be disputable. Reconsidering the quarter of birth instrument, positive monotonicity appears plausible in the U.S. context at a first glance. Arguably, among students entering school in the same year, those who are born in an earlier quarter can drop out after less years of completed education at the age of 16 when compulsory schooling ends than those born later, in particular after the end of the academic year. However, strategic postponement of school entry due to redshirting or unobserved school policies as discussed in Aliprantis (2012), Barua and Lang (2009), and Klein (2010) may reverse the relation of education and quarter of birth for some individuals such that defiers exist. Assumption 2.2 therefore needs to be scrutinized with similar care as Assumption 2.1.

The next key condition, Assumption 2.3, assumes the existence of compliers (type $T = c$) in the population.

Assumption 2.3 (First-stage)**.** $\Pr(D(1) > D(0)) > 0$

Under Assumption 2.1 and 2.2, $\Pr(D(1) > D(0)) > 0$ is equivalent to the existence of a first stage, $E(D|Z = 1) - E(D|Z = 0) > 0$ and thus corresponds to one of the two classical IV

assumptions. In our parametric model, this is satisfied if δ is positive and sufficiently large to shift the treatment decision at least for a subpopulation when switching from $z = 0$ to $z = 1$.

In seminal work, Vytlacil (2002) shows that Assumptions 2.1 – 2.3 correspond to a particular nonparametric IV model (2.1) with the following threshold crossing selection equation

$$D = 1(\mu(Z) \geq V), \tag{2.3}$$

where V is a scalar unobservable and $\mu(Z)$ is a nontrivial function of Z.

It is interesting to note that IV-based identification can also be obtained in structural models different from (2.2) and (2.2), which appear rather conventional. This concerns for instance the relation of Z and D, consider for instance the model provided in Hernan and Robins (2006):

$$Y = \phi(D, U), \quad D = \eta(V, U), \quad Z = \kappa(V), \quad U \perp V,$$

where $\phi(\cdot)$, $\eta(\cdot)$, $\kappa(\cdot)$ are unknown functions. Here, D is not affected by Z. However, the two variables are correlated through the unobservable V so that Z predicts D. As V and U are independent, Assumption 2.1 holds. Further examples can be found in Chalak and White (2011), who exhaustively discuss the structural relations under which a variable may serve as instrument Z in regression models.

2.2 Identification under a binary treatment and instrument

To demonstrate how Assumptions 2.1 – 2.3 permit identifying the LATE, LQTE, and the potential outcome distributions (including the means), we introduce some further notation that heavily borrows from Kitagawa (2009). Let $f(y, D = d|Z = z)$ denote the (observed) joint density of the observed outcome and $D = d$ conditional on $Z = z$ for $d, z \in \{1, 0\}$. Furthermore, denote by $f(y(d), T = t|Z = z)$ the unobserved joint density of the potential outcome and type t conditional on $Z = z$, where $t \in \{a, c, d, n\}$. In the absence of Assumptions 2.1 – 2.3, it follows from Table 2 that any observed joint density is a function of the potential outcomes of two different types conditional on Z, such that the subsequent relationships of observed and

unobserved joint densities hold for all y in the support of Y:

$$\begin{aligned} f(y, D=1|Z=1) &= f(y(1), T=c|Z=1) + f(y(1), T=a|Z=1), && (2.4)\\ f(y, D=1|Z=0) &= f(y(1), T=d|Z=0) + f(y(1), T=a|Z=0), && (2.5)\\ f(y, D=0|Z=1) &= f(y(0), T=d|Z=1) + f(y(0), T=n|Z=1), && (2.6)\\ f(y, D=0|Z=0) &= f(y(0), T=c|Z=0) + f(y(0), T=n|Z=0). && (2.7) \end{aligned}$$

When imposing Assumption 2.1, $f(y(d), T=t|Z=1) = f(y(d), T=t|Z=0) = f(y(d), T=t)$ for any type and treatment state, otherwise the potential treatment states and/or potential outcomes were not independent of the instrument. Under Assumption 2.2, $f(y(1), T=d)$ and $f(y(0), T=d)$ are equal to zero. Therefore, equations (2.4) to (2.7) simplify to

$$\begin{aligned} f(y, D=1|Z=1) &= f(y(1), T=c) + f(y(1), T=a), && (2.8)\\ f(y, D=1|Z=0) &= f(y(1), T=a), && (2.9)\\ f(y, D=0|Z=1) &= f(y(0), T=n), && (2.10)\\ f(y, D=0|Z=0) &= f(y(0), T=c) + f(y(0), T=n). && (2.11) \end{aligned}$$

where $f(y(0), T=c)$ and $f(y(1), T=c)$ are nonzero for at least some values $(y(0), y(1))$ in the support of $(Y(0), Y(1))$ by Assumption 2.3. Subtracting (2.9) from (2.8) and (2.10) from (2.11) yields the joint densities of the compliers under treatment and non-treatment:

$$f(y, D=1|Z=1) - f(y, D=1|Z=0) = f(y(1), T=c), \quad (2.12)$$

$$f(y, D=0|Z=0) - f(y, D=0|Z=1) = f(y(0), T=c). \quad (2.13)$$

To obtain the LATE, note that $\int f(y(d), T=c)dy = \pi_c$, where $\pi_c = \Pr(T=c)$ denotes the share of compliers in the population (and more generally, $\pi_t = \Pr(T=t)$ will henceforth denote

the share of type t). Therefore, π_c is identified by

$$\begin{aligned}\pi_c &= \int [f(y, D=1|Z=1) - f(y, D=1|Z=0)]dy \\ &= \Pr(D=1|Z=1) - \Pr(D=1|Z=0) = E(D|Z=1) - E(D|Z=0). \end{aligned} \tag{2.14}$$

Furthermore, $\int y[f(y(d), T=c)]dy = \int y[f(y(d)|T=c)]\pi_c dy = E[Y(d)|T=c] \cdot \pi_c$ implies that

$$\begin{aligned} & E[Y(1) - Y(0)|T=c] \cdot \pi_c \\ = & \int y\{[f(y, D=1|Z=1) - f(y, D=1|Z=0)] - [f(y, D=0|Z=0) - f(y, D=0|Z=1)]\}dy \\ = & \int y[f(y|Z=1) - f(y|Z=0)]dy = E(Y|Z=1) - E(Y|Z=0), \end{aligned}$$

which is the intention-to-treat effect (ITT). The latter generally deviates from the average treatment effect in the total population because it does not comprise the effects on the always and never takers, who do not react on the instrument. By scaling the ITT by the share of compliers we obtain the standard identification result for the LATE (denoted by Δ_c):

$$\frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} = E[Y(1) - Y(0)|T=c] = \Delta_c. \tag{2.15}$$

That is, the so-called *Wald estimand*, which in the binary treatment and instrument case corresponds to the probability limit of TSLS, identifies the LATE. It is worth noting that under one-sided noncompliance, the LATE simplifies to $\frac{E(Y|Z=1)-E(Y|Z=0)}{E(D|Z=1)}$ and coincides with the average treatment effect on the treated (ATT), $\Delta_{D=1}$, a parameter of major interest in the

treatment evaluation literature:

$$\begin{aligned}
\Delta_c &= E(Y(1) - Y(0)|D(1) = 1, D(0) = 0) \\
&= E(Y(1) - Y(0)|D(1) = 1) \\
&= E(Y(1) - Y(0)|D(1) = 1, Z = 1) \\
&= E(Y(1) - Y(0)|D = 1, Z = 1) \\
&= E(Y(1) - Y(0)|D = 1) \\
&= \Delta_{D=1}.
\end{aligned}$$

The second equality follows from $\Pr(D(0) = 1) = 0$ (one-sided non-compliance) such that $D(1) = 1$ implies $T = c$, the third from Assumption 2.1, the fourth from the definition of potential treatments, and the fifth from $\Pr(D(0) = 1) = 0 \Rightarrow \Pr(D = 1|Z = 0) = 0$ such that $D = 1 \Rightarrow D = 1, Z = 1$.

Also the density functions of the potential outcomes among compliers are identified, see Imbens and Rubin (1997). By (2.12) and (2.14)

$$\begin{aligned}
f(y(1)|T = c) &= \frac{f(y, D = 1|Z = 1) - f(y, D = 1|Z = 0)}{\Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0)} \\
&= \frac{f(y|D = 1, Z = 1) \cdot \Pr(D = 1|Z = 1) - f(y|D = 1, Z = 0) \cdot \Pr(D = 1|Z = 0)}{\Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0)}.
\end{aligned}$$

By (2.13) and (2.14)

$$\begin{aligned}
f(y(0)|T = c) &= \frac{f(y, D = 0|Z = 0) - f(y, D = 0|Z = 1)}{\Pr(D = 0|Z = 0) - \Pr(D = 0|Z = 1)} \\
&= \frac{f(y|D = 0, Z = 0) \cdot \Pr(D = 0|Z = 0) - f(y|D = 0, Z = 1) \cdot \Pr(D = 0|Z = 1)}{\Pr(D = 0|Z = 0) - \Pr(D = 0|Z = 1)}.
\end{aligned}$$

The mean potential outcomes among compliers correspond to the following expressions, see

also Imbens and Rubin (1997) and Abadie (2002):

$$\begin{aligned} E(Y(1)|T=c) &= \frac{\int y\{f(y,D=1|Z=1)-f(y,D=1|Z=0)\}dy}{\Pr(D=1|Z=1)-\Pr(D=1|Z=0)} \\ &= \frac{E(Y,D=1|Z=1)-E(Y,D=1|Z=0)}{E(D|Z=1)-E(D|Z=0)} \\ &= \frac{E(Y\cdot D|Z=1)-E(Y\cdot D|Z=0)}{E(D|Z=1)-E(D|Z=0)}. \end{aligned} \tag{2.16}$$

$$\begin{aligned} E(Y(0)|T=c) &= \frac{\int y\{f(y,D=0|Z=0)-f(y,D=0|Z=1)\}dy}{\Pr(D=0|Z=0)-\Pr(D=0|Z=1)} \\ &= \frac{E(Y,D=0|Z=0)-E(Y,D=0|Z=1)}{\Pr(D=0|Z=0)-\Pr(D=0|Z=1)} \\ &= \frac{E(Y\cdot(1-D)|Z=0)-E(Y\cdot(1-D)|Z=1)}{E(1-D|Z=0)-E(1-D|Z=1)} \\ &= \frac{E(Y\cdot(1-D)|Z=1)-E(Y\cdot(1-D)|Z=0)}{E(1-D|Z=1)-E(1-D|Z=0)}. \end{aligned} \tag{2.17}$$

$E(Y(1)|T=c)$ can be consistently estimated by a modified version of TSLS when using Z as instrument in a regression of $Y\cdot D$ on a constant and D, where the coefficient on the latter gives the estimate. Likewise, an estimate of $E(Y(0)|T=c)$ is obtained by a TSLS regression of $Y\cdot(1-D)$ on $(1-D)$.

As shown in Lemma 2.1 of Abadie (2002), the identification results (2.16) and (2.17) not only hold with respect to Y, but also for any function of the outcome, denoted by $h(y)$, with a finite first moment. As an important case, setting $h(y)=1(Y\leq y)$, with y being some value on the real line, allows identifying cumulative distribution functions (cdf) of potential outcomes:

$$\begin{aligned} F_{Y(1)|T=c}(y) &= \frac{E(1(Y\leq y)\cdot D|Z=1)-E(1(Y\leq y)\cdot D|Z=0)}{E(D|Z=1)-E(D|Z=0)}, \\ F_{Y(0)|T=c}(y) &= \frac{E(1(Y\leq y)\cdot(1-D)|Z=1)-E(1(Y\leq y)\cdot(1-D)|Z=0)}{E(1-D|Z=1)-E(1-D|Z=0)}, \\ F_{Y(1)|T=c}(y)-F_{Y(0)|T=c}(y) &= \frac{E(1(Y\leq y)|Z=1)-E(1(Y\leq y)|Z=0)}{E(D|Z=1)-E(D|Z=0)}. \end{aligned} \tag{2.18}$$

Estimation is straightforward by TSLS when regressing $1(Y\leq y)\cdot D$ on D and $1(Y\leq y)\cdot(1-D)$ on $(1-D)$, respectively.

Finally, quantiles of the potential outcomes of compliers are obtained by inverting the cdfs:

$$Q_{Y(d)|T=c}(\tau) = [\inf_y \Pr(Y(d) \leq y | T = c) \geq \tau] = F^{-1}_{Y(d)|T=c}(\tau), \tag{2.19}$$

where $\tau \in (0, 1)$ is the rank in the potential outcome distribution under $D = d$. This allows defining the local quantile treatment effect (LQTE) at the τ^{th} quantile, which corresponds to

$$\Delta_c(\tau) = Q_{Y(1)|T=c}(\tau) - Q_{Y(0)|T=c}(\tau). \tag{2.20}$$

Estimation can be performed by inverting the empirical potential outcome cdfs. Under standard regularity conditions the resulting estimators are consistent and asymptotically normal if the densities of the potential outcomes among compliers are positive at y: $f(y(d)|T = c) > 0$ for $d \in \{0, 1\}$.

2.3 Multivalued instruments and marginal treatment effects

In this section we consider extensions to setups with nonbinary instruments while maintaining the assumption that the treatment is binary.

First, if the instrument is multivalued one can identify a LATE with respect to any pair of values (z'', z') satisfying Assumptions 2.1 – 2.3. Instead of identifying many pairwise effects, we might be interested in the effect for the largest possible complier population. If we define monotonicity with respect to the treatment propensity score $p(z) = \Pr(D = 1|Z = z)$, this can be achieved by identifying the LATE with respect to the two instrument values that minimize and maximize $p(z)$, $(z_{\min}, z_{\max})$:

$$\Delta_c(p(z_{\min}), p(z_{\max})) = \frac{E(Y|p(Z) = p(z_{\max})) - E(Y|p(Z) = p(z_{\min}))}{E(D|p(Z) = p(z_{\max})) - E(D|p(Z) = p(z_{\min}))}.$$

If Z is multidimensional, the different elements in Z can straightforwardly be collapsed into a single instrument by using $p(Z)$.

If the instrument(s) is/are continuous, it is possible to identify a continuum of treatment

effects. This has been outlined in Heckman and Vytlacil (2001b) and Heckman and Vytlacil (2005), who call the resulting parameter based on an infinitesimal change in the instrument the marginal treatment effect (MTE). The latter is defined as average treatment effect conditional on V, the unobserved term in the treatment model (2.1):

$$\Delta(v) = E(Y(1) - Y(0)|V = v).$$

Assume that V represents the (unobserved) cost or disutility of treatment. The MTE can then be interpreted as average effect among persons who would be indifferent between treatment or not if exogenously assigned a value of Z, say z, such that $\mu(z) = v$, which follows from the treatment model representation $D = 1(\mu(Z) \geq V)$, see (2.3). Any LATE (and any other average treatment effect) can be expressed as a (density-)weighted average of MTEs. Note that for any (z'', z') such that $p(z'') > p(z')$, a complier is someone satisfying $D(z'') = 1(\mu(z'') \geq V) = 1$ and $D(z') = 1(\mu(z') \geq V) = 0$. Put differently, compliers $c(z'', z')$ are characterized by $v' < V \leq v''$ so that $D(z'') = 1$ and $D(z') = 0$ holds. Therefore, the LATE for $T = c(z'', z')$ is defined as

$$\begin{aligned} E(Y(1) - Y(0)|T = c(z'', z')) &= E(Y(1) - Y(0)|D(z'') = 1, D(z') = 0) \\ &= E(Y(1) - Y(0)|v' < V \leq v'') \\ &= \Delta_c(v'', v') = \frac{1}{F_V(v'') - F_V(v')} \int_{v'}^{v''} \Delta(v) dF_V(v). \end{aligned}$$

V can be normalized so that the normalization (denoted by $\bar{V}$) satisfies $\bar{V} \sim$ Uniform$[0, 1]$. Therefore, the normalization corresponds to the cdf: $\bar{V} = F_V$. This normalization is innocuous given our assumptions, because if $D = 1(\mu(Z) \geq V)$, then by applying a probability transformation, the model can be reparametrized so that $D = 1(\eta(Z) \geq \bar{V})$, with $\eta(Z) = F_V(\mu(Z))$. It follows that

$$\Delta_c(\bar{v}'', \bar{v}') = \frac{1}{\bar{v}'' - \bar{v}'} \int_{\bar{v}'}^{\bar{v}''} \Delta(\bar{v}) d\bar{v}.$$

The MTE can be identified by the fact that $\bar{v}'' = F_V(v'') = \Pr(D(z'') = 1) = \Pr(D = 1|Z =$

$z'') = p(z'')$ and equivalently, $\bar{v}' = p(z')$. Therefore, the MTE is recovered pointwise by the derivative of the conditional expectation of Y with respect to $p(Z)$:

$$\Delta(\bar{V} = p(z)) = \frac{\partial E(Y|p(Z) = p(z))}{\partial p(z)}.$$

This follows from the fact that

$$\begin{aligned} E(Y|p(Z) = p(z)) &= E(Y(0)|p(Z) = p(z)) + E(Y(1) - Y(0)|p(Z) = p(z), D = 1) \cdot p(z) \\ &= E(Y(0)) + E(Y(1) - Y(0)|p(z) \geq \bar{V}) \cdot p(z) \\ &= E(Y(0)) + \int_0^{p(z)} \Delta(\bar{v}) d\bar{v}, \end{aligned}$$

such that the first derivative yields the parameter of interest. Heckman and Vytlacil (1999) coined the term local IV (LIV) for $\Delta(\bar{V} = p)$, a parameter even 'more local' than the conditional LATE $\Delta_c(\bar{v}'', \bar{v}')$ based on a quantifiable difference between $\bar{v}''$ and $\bar{v}'$. Note, however, that the conditional LATE is equivalent to the LIV for $\bar{v}'' - \bar{v}'$ infinitesimally small.

Using similar arguments, Carneiro and Lee (2009) extend these ideas to the identification of the QTE analogs of the MTE, the marginal quantile treatment effects (MQTE):

$$\Delta(\tau|\bar{V} = p(z)) \equiv Q_{Y_1}(\tau|\bar{V} = p(z)) - Q_{Y_0}(\tau|\bar{V} = p(z)).$$

$Q_{Y_1}(\tau|\bar{V} = p(z))$ and $Q_{Y_0}(\tau|\bar{V} = p(z))$ are identified as the inverses of

$$\begin{aligned} F_{Y(1)}(y|\bar{V} = p(z)) &= F_Y(y|P(Z) = p(z), D = 1) + p(z)\frac{\partial F_Y(y|P(Z) = p(z), D = 1)}{\partial p}, \\ F_{Y(0)}(y|\bar{V} = p(z)) &= F_Y(y|P(Z) = p(z), D = 0) - (1 - p(z))\frac{\partial F_Y(y|P(Z) = p(z), D = 0)}{\partial p}. \end{aligned}$$

2.4 Multivalued treatments

In contrast to extensions of the standard LATE framework to multivalued instruments as considered in Section 2.3, generalizing binary to nonbinary treatments is not straightforward. To

illustrate this point, consider a setup with a single binary instrument $Z \in \{0,1\}$ and an ordered discrete treatment $D \in \{0,1,...,J\}$, where $J+1$ is the number of possible treatment doses. We cannot identify causal effects for single compliance types at specific treatment values, e.g. for those increasing the treatment from 1 to 2 when the instrument switches from 0 to 1. However, it is possible to identify a weighted average of causal effects of unit increases in the treatment, $Y(j) - Y(j-1)$, $j \in \{1,...,J\}$. Specifically, Angrist and Imbens (1995) show that if $\Pr(D(1) \geq j > D(0)) > 0$ for at least one value j such that compliers exist at some margin of the treatment, we have that

$$\frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} = \sum_{j=1}^{J} w_j \cdot E(Y(j) - Y(j-1)|D(1) \geq j > D(0)), \tag{2.21}$$

where

$$w_j = \frac{\Pr(D(1) \geq j > D(0))}{\sum_{j=1}^{J} \Pr(D(1) \geq j > D(0))},$$

implying that $0 \leq w_j \leq 1$ and $\sum_{j=1}^{J} w_j = 1$. Therefore, the Wald estimand equals a weighted average of per-unit treatment effects, where, unfortunately, the weights cannot be identified. Angrist and Imbens (1995) show that similar results hold in setups with multiple instruments and covariates. It is important to note that while this strategy yields weighted LATEs, it cannot be applied to identify LQTEs as their identification hinges on separately identifying and subsequently inverting marginal distributions of potential outcomes.

Several contributions discuss identification when treatment values cannot be ordered. Behaghel et al. (2013) consider multiple unordered treatments that are mutually exclusive, which is equivalent to the case of a single treatment with multiple, albeit unordered values. They demonstrate under Assumption 2.1 and a specific monotonicity assumption tailored to the investigated case of a three-valued treatment and instrument ($D, Z \in \{0,1,2\}$) that LATEs among the two complier populations $c_1 : D(1) = 1, D(0) = 0$ and $c_2 : D(2) = 2, D(0) = 0$ are identified. Heckman and Pinto (2015) consider an unordered monotonicity assumption that requires for any specific value of the unordered treatment that if some subjects move into (out of) the

respective value when the instrument is switched, then no subjects can at the same time move out of (into) that value. Hull (2015) imposes conditional IV validity in the spirit of Assumption 3.1 and shows under a modified monotonicity assumption for a three-valued treatment that LATEs can be obtained even from a binary instrument if (i) compliance is heterogeneous and (ii) LATEs are homogeneous in observables X. Lee and Salanie (2015) discuss identification under the conditions that any treatment value is a measurable function of some threshold-crossing models and sufficiently many continuous instruments are available, but require no classical monotonicity assumption.

3 Treatment evaluation with covariates

We subsequently discuss IV-based treatment evaluation in the presence of covariates. Section 3.1 presents the identifying assumptions, while Sections 3.2, 3.3, 3.4 consider the evaluation of local, quantile, and marginal treatment effects, respectively. Section 3.5 shows that quite general functionals rather than merely effects can be identified for compliers.

3.1 Identifying assumptions

It may not appear credible that an instrument satisfies Assumptions 2.1 – 2.3 unconditionally, i.e. without controlling for further covariates. As an example, consider the study of Card (1995), who evaluates the returns to college education using the U.S. National Longitudinal Survey of Young Men. Geographic proximity to college serves as instrument for the potentially endogenous decision of going to college. Proximity should induce some individuals to strive for a college degree who would otherwise not, for instance due to costs associated with not living at home. However, the instrument might be correlated with factors like local labor market conditions or family background which might be related to the earnings outcome, implying a violation of Assumption 2.1. For these reasons, Card (1995) includes a range of control variables in his estimations, including parents' education, ethnicity, urbanity, and geographic region.

We subsequently reconsider the binary instrument and treatment case of Section 2.1, but now impose conditional IV assumptions (see for instance Abadie (2003)), which imply that the IV assumptions only hold when controlling for a vector of observed covariates denoted by X.

Assumption 3.1 (Conditional independence). $Z \perp (D(1), D(0), Y(1), Y(0))|X$

Assumption 3.2 (Monotonicity). $\Pr(D(1) \geq D(0)|X) = 1$

Assumption 3.3 (First-stage). $\Pr(D(1) > D(0)|X) > 0$

Assumption 3.1 is weaker than Assumption 2.1, because independence now only holds among units with the same values of X, implying that Z is as good as randomly assigned given X. Assumption 3.2 requires that defiers do not exist for every value of X. Theoretically, one could construct cases where defiers exist unconditionally (such that $\Pr(D(1) \geq D(0)) = 1$ as stated in Assumption 2.2 does not hold), but not after conditioning on X, for instance if Z affected X positively and X affected D (sufficiently strongly) negatively. Assumption 3.3 implies that compliers exist for every value of X in its support, which is stronger than Assumption 2.3. $\Pr(D(1) > D(0)|X) > 0$ is required for identifying the conditional LATE or LQTE, see Sections 3.2 and 3.3 almost everywhere, while $\Pr(D(1) > D(0)) > 0$ suffices if one is only interested in the (unconditional) LATE and LQTE.

Assumption 3.4 (Common support). $0 < \Pr(Z = 1|X) < 1$

Assumption 3.4 is a common support restriction requiring that no value of X perfectly predicts (non-)assignment to the instrument. If it was not satisfied, no comparable units (in terms of X) across instrument states $Z = 1$ and $Z = 0$ would exist for some values of X so that identification would break down at these values.

Similar to (2.1), we briefly consider a general IV model that now includes X to further elucidate the implications of the assumptions:

$$\begin{aligned} Y &= \phi(D, X, U), \quad Y(d) = \phi(d, X, U), \\ D &= \delta(Z, X, V), \quad D(z) = \delta(z, X, V). \end{aligned} \tag{3.1}$$

Assumption 3.1 implies that $(U, V) \perp Z|X$. Furthermore, under Assumptions 3.2 – 3.3 , D can also be represented as $D = 1(\mu(Z, X) \geq V)$; see Vytlacil (2002).

3.2 LATE

Using analogous arguments as in Section 2, the conditional LATE given $X = x$ is identified under Assumptions 3.1 – 3.4 by

$$E(Y(1) - Y(0)|T = c, X = x) = \Delta_c(x) = \frac{E(Y|Z = 1, X = x) - E(Y|Z = 0, X = x)}{E(D|Z = 1, X = x) - E(D|Z = 0, X = x)}, \quad (3.2)$$

see for instance Heckman (1997). Nonparametric estimation of $\Delta_c(x)$ suffers from the curse of dimensionality when X is high dimensional. To overcome this problem, one may either impose parametric restrictions on the conditional means $E(Y|Z = z, X = x)$ and $E(D|Z = z, X = x)$ for $z \in \{0, 1\}$, see Tan (2006), or employ a semiparametric approach based on the weighting result by Abadie (2003) to construct weighted least squares estimates, see Section 3.5.

While identification of the conditional LATE allows investigating effect heterogeneity with respect to observable covariates, the (unconditional) LATE is frequently the main parameter of interest also under the conditional IV assumptions. It is obtained as a weighted average of conditional LATEs among compliers, i.e. by integrating over the distribution X given $T = c$:

$$\Delta_c = \int \Delta_c(x) dF_{X|T=c}(x).$$

Frölich (2007) shows that the LATE can also be represented in the following way:

$$\Delta_c = \frac{\int \{E(Y|Z = 1, X = x) - E(Y|Z = 0, X = x)\} dF_X(x)}{\int \{E(D|Z = 1, X = x) - E(D|Z = 0, X = x)\} dF_X(x)}. \quad (3.3)$$

To see why (3.3) holds, note that $\Delta_c = \int \Delta_c(x) dF_{X|T=c}(x)$. By Bayes' theorem, $dF_{x|T=c} = \Pr(T = c|X = x)/\pi_c dF_X(x)$ so that $\Delta_c = \int \Delta_c(x) \Pr(T = c|X = x)/\pi_c dF_X(x)$. Finally,

plugging $\Delta_c(x)$ into the last equation yields

$$\begin{aligned}\Delta_c &= \frac{\int\{E(Y|Z=1,X=x)-E(Y|Z=0,X=x)\}dF_X(x)}{\pi_c} \\ &= \frac{\int\{E(Y|Z=1,X=x)-E(Y|Z=0,X=x)\}dF_X(x)}{\int \Pr(T=c|X=x)dF_X(x)} \\ &= \frac{\int\{E(Y|Z=1,X=x)-E(Y|Z=0,X=x)\}dF_X(x)}{\int\{E(D|Z=1,X=x)-E(D|Z=0,X=x)\}dF_X(x)}.\end{aligned}$$

By noting that $\int E(Y|Z=1,X=x)dF_X(x) = \int(1/\pi(x))E(Y\cdot Z|X=x)dF_X(x) = E(Y\cdot Z/\pi(X))$, where $\pi(X)=\Pr(Z=1|X)$ is the instrument propensity score, a weighting-based expression is also obtained, see Tan (2006) and Frölich (2007):

$$\Delta_c = \frac{E[Y\cdot Z/\pi(X) - Y(1-Z)/(1-\pi(X))]}{E[D\cdot Z/\pi(X) - D(1-Z)/(1-\pi(X))]}. \tag{3.4}$$

Finally, making use of a result of Rosenbaum and Rubin (1983) showing that controlling for the propensity score is in terms of identification as good as controlling for X when evaluating average effects, a third representation of the LATE is given by

$$\Delta_c = \frac{\int\{E(Y|Z=1,\pi(X)=p)-E(Y|Z=0,\pi(X)=p)\}dF_\pi(p)}{\int\{E(D|Z=1,\pi(X)=p)-E(D|Z=0,\pi(X)=p)\}dF_\pi(p)}, \tag{3.5}$$

which has the practical advantage that $\pi(X)$ is one-dimensional, no matter of which dimension X is. This implies that the LATE can be estimated as the ratio of two propensity score matching estimators with Z being the 'treatment' and either Y (numerator) or D (denominator) being the 'outcome'.

Several analog estimators have been proposed based on (3.3), (3.4), and (3.5). Frölich (2007) analyzes nonparametric matching- and (local polynomial and series) regression-based estimation of (3.3), while Belloni et al. (2014) derive the properties of regression-based estimators of (3.3) in data-rich environments. Donald et al. (2014b) and Donald et al. (2014a) propose nonparametric inverse probability weighted estimators of (3.4), using series logit and local polynomial regression-based estimation, respectively, of the instrument propensity score. All of these estimators are $\sqrt{n}$-consistent and asymptotically normal under appropriate regularity

conditions. The reason is that fully nonparametric estimation of unconditional LATE involves averaging over conditional LATEs and does therefore not give rise to the curse of dimensionality. Parametric estimation strategies for the unconditional LATE are outlined in Tan (2006) and Uysal (2011), who both propose estimators that rely on parametric models for the propensity scores and conditional expectations. To guard against misspecification, they consider so-called doubly-robust (DR) estimators. DR estimators are consistent if either the propensity score, the conditional expectations, or both are correctly specified. Finally, Hong and Nekipelov (2010) provide general semiparametric efficiency results for the estimation of nonlinear LATE models.

When the IV assumptions hold conditionally on X, the LATE among all compliers is different from the local average treatment effect among treated compliers (LATT), as the distribution of X generally differs across treatment states. By appropriate reweighting of the previous identification results, also the LATT is identified. For instance, by weighting observations in expression (3.4) by $\pi(X)/\Pr(Z=1)$ one obtains the LATT, see Donald et al. (2014b):

$$\Delta_{c,D=1} = \frac{\pi(X)\cdot(E[Y\cdot Z/\pi(X) - Y(1-Z)/(1-\pi(X))])}{\pi(X)\cdot(E[D\cdot Z/\pi(X) - D(1-Z)/(1-\pi(X))])}. \tag{3.6}$$

Note that in the case of one-sided non-compliance given X, $\Pr(D(0)=1|X)=0$, the LATE does not correspond to the ATT under Assumptions 3.1 – 3.4 (in contrast to Assumptions 2.1 – 2.3). Frölich and Melly (2013a) show that in this case, the ATT is identified by

$$\begin{aligned}\Delta_{D=1} &= \frac{E(Y) - \int E(Y|Z=0,X=x)dF_X(x)}{\Pr(D=1)} \\ &= \frac{1}{\Pr(D=1)} E\left[Y\cdot\left(D-(1-D)\cdot\frac{\pi(X)-Z}{1-\pi(X)}\right)\right].\end{aligned}$$

3.3 LQTE

As for the LATE, one may define either conditional (given X) or unconditional LQTEs in the presence of covariates. This distinction is important because of the definition of quantiles. Suppose that we are interested in the relationship between education and wages. The unconditional 0.9 quantile of the wage distribution refers to high wage workers who typically have many years

of schooling, whereas the 0.9 quantile of the wage distribution conditional on schooling refers to the high wage earners within an education class who will not necessarily be high overall earners. See also Frölich and Melly (2013b), who provide a more detailed discussion about the difference between conditional and unconditional LQTEs. Abadie et al. (2002) consider estimation of the conditional LQTE. Assuming that the conditional quantile function for the compliers satisfies

$$Q_{Y|D,X,T=c}(\tau) = \alpha_c(\tau)D + X'\beta_c(\tau), \tag{3.7}$$

they show that conditional LQTE, $\Delta_c(\tau|x)$, is identified by $\alpha_c(\tau)$, the coefficient on D in the following weighted quantile regression objective function:

$$(\alpha_c(\tau), \beta_c(\tau)) = \arg\min_{a,b} E[\kappa \cdot \rho_\tau(Y - aD - X'b)]. \tag{3.8}$$

κ, which is defined in Section 3.5 below, is a weighting function that allows identifying functionals for compliers. Note that among the population of compliers, outcome comparisons by D conditional on X as in (3.7) have a causal interpretation, which follows from Assumption 3.1 and the fact that compliers satisfy $D = Z$:

$$\begin{aligned} Z \perp (D(1), D(0), Y(1), Y(0))|X \quad &\Rightarrow \quad Z \perp (Y(1), Y(0))|X, T = c \\ &\Rightarrow \quad D \perp Y(1), Y(0)|X, T = c. \end{aligned}$$

Although the population objective function (3.8) is globally convex, its sample counterpart is typically not because κ is negative when $D \neq Z$, see the discussion in Section 3.5. Abadie et al. (2002) therefore suggest replacing the κ-weights by their projections on (Y, D, X), which are guaranteed to be positive. Their estimation strategy consists of two steps: (i) nonparametric power series estimation of the weights and (ii) a weighted quantile regression using the estimated weights from the first step. Under appropriate regularity conditions, the resulting estimators $\hat{\alpha}_c(\tau)$ and $\hat{\beta}_c(\tau)$ are $\sqrt{n}$-consistent and asymptotically normal, because the outcome equation is parametric. As for the conditional LATE, conditional LQTE cannot be estimated at the

$\sqrt{n}$-rate without parametric assumptions.

Concerning unconditional LQTE estimation when controlling for covariates, first note that the unconditional complier cdf is, in analogy to (2.18) combined with (3.3), identified as

$$\begin{aligned} F_{Y(1)|T=c}(y) &= \frac{\int\{E[1(Y \leq y) \cdot D|Z=1, X=x] - E[1(Y \leq y) \cdot D|Z=0, X=x]\} dF_X(x)}{\int\{E(D|Z=1, X=x) - E(D|Z=0, X=x)\} dF_X(x)} \\ &= \frac{E(\kappa_{FM} \cdot 1(Y \leq y) \cdot D)}{E(\kappa_{FM} \cdot D)}, \end{aligned} \tag{3.9}$$

see (Frölich and Melly, 2013b), where

$$\kappa_{\text{FM}} = \frac{Z - \pi(X)}{\pi(X) \cdot (1 - \pi(X))} \cdot (2D - 1). \tag{3.10}$$

An analogous result holds for $F_{Y(0)|T=c}(y)$ by replacing D with $1-D$, such that the unconditional LQTE is given by

$$\Delta_c(\tau) = F^{-1}_{Y(1)|T=c}(\tau) - F^{-1}_{Y(0)|T=c}(\tau).$$

Alternatively, the unconditional QTE can be identified from the following weighted quantile regression problem:

$$(\alpha_c(\tau), \beta_c(\tau)) = \arg\min_{a,b} E[\kappa_{\text{FM}} \cdot \rho_\tau(Y - aD - b)]. \tag{3.11}$$

Finally, Frölich and Melly (2013a) show that under one-sided noncompliance, the quantile treatment effect on the treated is given by

$$\Delta_{D=1}(\tau) = Q_{Y|D=1}(\tau) - F^{-1}_{Y(0)|D=1}(\tau),$$

where

$$
\begin{aligned}
F_{Y(0)|D=1}(\tau) &= \frac{\int E[1(Y \leq q)|Z=0, X=x]dF_X(x) + E[1(Y \leq q) \cdot (D-1)]}{\Pr(D=1)} \\
&= \frac{1}{\Pr(D=1)} E\left[1(Y \leq q) \cdot (1-D) \cdot \frac{\pi(X) - Z}{1-\pi(X)}\right].
\end{aligned}
$$

Representations (3.9), (3.10), and (3.11) suggest estimators based on the respective sample analogs. Belloni et al. (2014) consider regression-based estimators of (3.9) in data-rich environments. Hsu et al. (2015) derive uniformly consistent and asymptotically Gaussian estimators of (3.10) using series logit regression for propensity score estimation. Frölich and Melly (2013b) estimate (3.11) using local polynomial regression for propensity score estimation.

3.4 Marginal treatment effects

In the presence of covariates, the marginal treatment effect given X, $\Delta(v,x) = E(Y(1) - Y(0)|V=v, X=x)$, is identified by LIV,

$$
\Delta(\bar{V} = p(z,x)) = \frac{\partial E(Y|p(Z,X) = p(z,x))}{\partial p(z,x)},
$$

with $p(z,x) = \Pr(D=1|Z=z, X=x)$, given that Assumptions 3.1 – 3.3 hold for all values of $p(Z,X)$ of interest. Assumption 3.4 adapted to the continuous instrument $p(Z,X)$ implies that the MTE is only identified over the common support of $p(Z,X)$ across all values of X. This limits the feasibility of nonparametric MTE evaluation in practice, in particular if X is high dimensional and Z is not excessively strong or rich in support. We refer to Cornelissen et al. (2016) for an introduction and overview of different methods for estimating MTE with covariates.

As discussed in Carneiro et al. (2011), identifying power is increased if Assumption 3.1 is replaced by the following condition:

Assumption 3.5. $(Z,X) \perp (D(z), D(z'), Y(1), Y(0))$ *for* z, z' *in the support of* Z

Note that $z = 1, z' = 0$ in the binary instrument case. This restriction imposes the inde-

pendence of X and unobservables affecting the treatment or the outcome, which is therefore substantially stronger than Assumption 3.1. While observed characteristics X as for instance education or age are allowed to confound Z and D, Y they are not allowed to be associated with unobservables as for instance motivation or ability that affect D, Y. If Assumption 3.5 is nevertheless imposed, the MTE is (similarly as under Assumptions 2.1 – 2.3 and in the absence of X) identified over the unconditional support of $p(Z, X)$.

Identification of MQTE in the presence of covariates follows from the same arguments as discussed in Section 2.3 conditional on X. Carneiro and Lee (2009) propose a semiparametric estimation approach which relies on additive separability of the structural functions determining potential outcomes and derive its asymptotic properties. In contrast, Yu (2014) proposes a semiparametric estimation strategy that does not rely on separability of the structural functions. He derives the corresponding weak limits and shows validity of the bootstrap for inference.

3.5 General functionals

Abadie (2003) shows that under Assumptions 3.1 – 3.4 , it is possible to identify a broad class of functionals for the compliers, rather then merely treatment effects. For any real function $g(Y, D, X)$ with a finite first moment and weighting functions

$$\begin{aligned}
\kappa_{(0)} &\equiv (1-D)\cdot\frac{(1-Z)-(1-\pi(X))}{(1-\pi(X))\cdot\pi(X)},\\
\kappa_{(1)} &\equiv D\cdot\frac{Z-\pi(X))}{(1-\pi(X))\cdot\pi(X)},\\
\kappa &\equiv \kappa_{(0)}\cdot(1-\pi(X))+\kappa_{(1)}\cdot\pi(X)=1-\frac{D\cdot(1-Z)}{1-\pi(X)}-\frac{(1-D)\cdot Z}{\pi(X)},
\end{aligned}$$

it holds that

$$\begin{aligned}
E(g(Y,D,X)|T=c) &= \frac{E(\kappa\cdot g(Y,D,X))}{E(\kappa)},\\
E(g(Y(0),X)|T=c) &= \frac{E(\kappa_{(0)}\cdot g(Y,X))}{E(\kappa)},\\
E(g(Y(1),X)|T=c) &= \frac{E(\kappa_{(1)}\cdot g(Y,X))}{E(\kappa)}.
\end{aligned}$$

In words, the weighting functions κ, $\kappa_{(1)}$, and $\kappa_{(0)}$ allow identifying functions (e.g. conditional expectations and regression functions) for compliers, for compliers under treatment, and for compliers under non-treatment, respectively. To see this, note for instance for κ that by the law of iterated expectations,

$$\begin{aligned} E(\kappa) &= E(1 - \Pr(D = 1|Z = 0, X) - \Pr(D = 0|Z = 1, X)) \\ &= E(1 - \Pr(T = a|X) - \Pr(T = n|X)) \\ &= E(\Pr(T = c|X)) = \pi_c, \end{aligned}$$

implying that $E(\kappa \cdot g(Y, D, X))/E(\kappa) = E(g(Y, D, X)|T = c)$. However, κ does not produce proper weights since it takes negative values when D differs from Z.

Section 3.3 has presented an application of this general weighting result to evaluate the conditional LQTE. As a further application, consider the linear outcome model, $Y = X'\alpha + \beta D + U$, with $E[U|X, D] = 0$ and α, β denoting the coefficients on the covariates and the treatment, respectively. The optimization problem is

$$(\alpha_c, \beta_c) = \arg\min_{a,b} E((Y - X'a - bD)^2|T = c) = \arg\min_{a,b} E(\kappa \cdot (Y - X'a - bD)^2).$$

Note that division by $E(\kappa)$ is not required as it does not affect the minimization problem. β_c gives the conditional LATE $\Delta_c(x)$, which in our linear model also corresponds to the (unconditional) LATE Δ_c, as well as the treatment effect in the entire population. In contrast to TSLS, this approach does not require specifying a first stage equation about the relationship of D, Z, and X, but instead relies on a model for the instrument propensity score $\pi(X)$. Abadie (2003) provides conditions under which two-step estimators based on the weighting functions κ, $\kappa_{(1)}$, and $\kappa_{(0)}$ are consistent and asymptotically normal.

4 Some extensions

The following sections briefly discuss extensions of the IV framework to more complex identification problems. Section 4.1 presents approaches to LATE evaluation under outcome attrition, outcome non-response, or sample selection. Section 4.2 discusses methods dealing with measurement errors in the treatment or the instrument. Section 4.3 considers identifying the effects of dynamic, i.e. sequentially assigned, or multiple treatments. Section 4.4 is concerned with disentangling the (total) LATE into various causal mechanisms or direct and indirect effects.

4.1 LATE evaluation under outcome attrition and sample selection

In addition to treatment endogeneity, treatment evaluation is frequently complicated by selective attrition bias in the outcome, for instance due to drop-out bias in a follow-up survey in which the outcome is measured after a randomized trial or due to sample selection, e.g. when the wage outcome is only observed for the working. Outcome non-response is frequently modelled by a so-called missing-at-random (MAR) restriction, which assumes conditional independence of attrition and outcomes given observed variables (e.g. Z, D, X), see for instance Rubin (1976) and Little and Rubin (1987). An alternative to MAR which is particularly tailored to the LATE framework is the so-called latent ignorability (LI) assumption of Frangakis and Rubin (1999), which requires outcome non-response to be independent of the potential outcomes conditional on the compliance type. Furthermore, MAR and LI might be combined such that independence is assumed conditional on both observed characteristics and compliance types, see for instance Mealli et al. (2004):

$$Y \perp R | Z, T, X,$$

where R is a binary indicator for observing outcome Y. Note that this condition is equivalent to $Y \perp R | Z, D, T, X$ as Z and T perfectly determine D. Frölich and Huber (2014b) extend LATE identification under MAR and both MAR and LI to dynamic non-response models with multiple outcome periods.

A shortcoming of LI (and MAR) is that outcome non-response must not be related in a very

general way to unobservables affecting the outcome, because the compliance type is essentially assumed to serve as sufficient statistic for the association between response and unobservables, at least conditional on observed variables. So-called non-ignorable non-response models do not impose such restrictions on the relation of R and, for instance, U in (3.1). However, without a second instrument for R, the LATE is only identified under tight structural assumptions, see for instance Zhang et al. (2009) and Frumento et al. (2012). In contrast, Fricke et al. (2015) discuss nonparametric LATE identification when a continuous instrument for non-response is available in addition to the binary instrument for the treatment and present an application in which either instrument is independently randomized from each other. Chen and Flores (2015) do not consider instruments, LI, or MAR with respect to response, but partially identify the LATE based on imposing monotonicity of R in D among compliers as well as a particular order of mean potential outcomes under specific treatments across various subpopulations defined in terms of compliance and response.

4.2 Measurement error in the treatment or instrument

Ura (2016) discusses LATE evaluation when the treatment is measured with error, i.e., misclassified. While point identification is generally lost, the study provides upper and lower bounds when Assumptions 2.1 – 2.3 are satisfied with respect to the true treatment. Ura (2016) clarifies that the Wald estimand generally lies outside the identified set and is only included in the latter if the conditions (6.1) in Section 6.1 are satisfied with respect to the mismeasured treatment. In contrast, Chalak (2016) considers measurement error in the instrument rather than the treatment. Denoting by W, Z the mismeasured and true instrument, respectively, W is assumed to be mean independent of Y and D given Z and to satisfy an exclusion restriction, while monotonicity is not imposed. In the binary instrument case and under the satisfaction of Assumptions 2.1 – 2.3, W identifies the same LATE that would have been recovered under Z. For more general settings with multiple treatment and/or instrument values, Chalak (2016) shows that the Wald and LIV estimands using W identify weighted averages of LATEs or MTEs and discusses necessary and sufficient conditions for the weights being nonnegative.

4.3 Dynamic and multiple treatments

Rather then evaluating the effects of single treatments, one might be interested in the impact of several sequentially assigned (i.e. dynamic) treatments that take place at various points in time. Consider for instance the effectiveness of sequences of active labor market policies like a job application training, which is followed by an IT course and a subsidized employment program. This sequence could be compared to non-participation in any program or a different sequence of interventions. Such a dynamic treatment framework generally requires multiple instruments for each of the treatments and specific multi-period monotonicity conditions. More formally, consider a set up with two treatment periods and let D_1, D_2 denote the first and second binary treatment, respectively, and $Y(d_1, d_2)$ the potential outcome potential outcome now defined in terms of two treatment interventions (with $d_1, d_2 \in \{1, 0\}$). Furthermore, let T_1 and T_2 denote the compliance types defined in terms of the reaction of D_1 to the first instrument Z_1 and of D_2 to the second instrument Z_2. Miquel (2002) discusses various conditions under which dynamic LATEs are identified for specific types defined in terms of first- and second-period compliance, respectively. Among others, she considers the identification among compliers w.r.t. either instrument:

$$E[Y(d_1, d_2') - Y(d_1'', d_2''')|T_1 = c, T_2 = c] \text{ for } d_1, d_2', d_1'', d_2''' \in \{1, 0\}.$$

Miquel (2002) also shows that if only one instrument is available for both treatment periods, only the effects of particular sequences can be identified under specific assumptions for individuals that are always or never takers in the first treatment and compliers in the second one or vice versa.

If various treatments are not assigned sequentially, but rather at the same point of time such that participation in the first treatment does not affect participation in the second one, we are in a multiple treatment framework. At a first glance, the simultaneous availability of several binary treatments (e.g. alternative active labor market policies) constitutes a similar evaluation problem like a treatment with multiple ordered values as discussed in Section 2.4. However, if no natural ordering between the various treatments arises, the monotonicity assumption is likely

violated and even under the satisfaction of monotonicity, the weighted effect given in (2.21) is hard to interpret. Furthermore, in the multiple treatment case, one might be interested in the effect of assigning several treatments at the same time. Therefore, one generally requires distinct instruments for each treatment. Blackwell (2015) considers LATE identification of separate and joint effects of two treatments in various subpopulations defined upon compliance with either of the binary instruments, namely: $E[Y(1,1) - Y(0,0)|T_1 = c, T_2 = c]$, $E[Y(1,0) - Y(0,0)|T_1 = c, T_2 \in \{c, n\}]$, $E[Y(1,1) - Y(0,1)|T_1 = c, T_2 \in \{c, a\}]$.

4.4 Direct and indirect effects (causal mechanisms)

As a further extension that is related to dynamic treatment effects, consider the problem of disentangling the total impact of a treatment into a direct effect and an indirect effect that operates via an intermediate variable (or so-called mediator) which also affects the outcome. That is, the interest lies in disentangling a treatment effect into various causal mechanisms, which may provide a better understanding of why specific treatments are effective or ineffective by opening the 'black box' of the total effect. As an example, consider the health effect of college attendance (D_1), which likely affects the employment state (D_2) which also influences the health outcome. Disentangling the direct effect of college attendance from its indirect effect operating via employment shows whether the health impact of college attendance is only driven via its impact on labor market participation, or also through other ("direct") channels, for instance college peers-induced adaption of health behaviour.

To formally define the effects of interest, let $D_2(d_1)$ denote the potential state of the second treatment as a function of the first. The standard notation for potential outcomes defined in terms of D_1 can then easily be linked to the notation appropriate to analysing causal mechanisms, namely: $Y(d_1) = Y(d_1, D_2(d_1))$, which makes explicit that D_1 might affect Y either directly or indirectly through its effect on D_2. Therefore, the LATE of the first treatment among compliers in the first treatment period can be expressed as

$$\Delta_{c1} = E[Y(1) - Y(0)|T_1 = c] = E[Y(1, D_2(1)) - Y(0, D_2(0))|T_1 = c], \tag{4.1}$$

and comprises both the direct and indirect effect of D_1 on Y.

The direct effect, denoted by $\theta_{c1}(d_1)$, is obtained by shutting down the indirect causal mechanism by fixing D_2 to its potential value under a particular d_1, while exogenously varying the first treatment D_1:

$$\theta_{c1}(d_1) = E[Y(1, D_2(d_1)) - Y(0, D_2(d_1))|T_1 = c], \qquad \text{for } d_1 \in \{0, 1\}. \tag{4.2}$$

The indirect effect among compliers, denoted by $\delta_{c1}(d_1)$, corresponds to the mean difference in outcomes when exogenously shifting D_2 to its potential values for $d_1 = 1$ and $d_1 = 0$, but keeping the first treatment fixed at $D_1 = d_1$:

$$\delta_{c1}(d_1) = E[Y(d_1, D_2(1)) - Y(d_1, D_2(0))|T_1 = c], \qquad \text{for } d_1 \in \{0, 1\}. \tag{4.3}$$

The LATE is the sum of the direct and indirect effects defined upon opposite states of d_1, which can be seen from adding and subtracting either $Y(0, D_2(1))$ or $Y(1, D_2(0))$ in (4.1):

$$\begin{aligned}
\Delta_{c1} &= E[Y(1, D_2(1)) - Y(0, D_2(0))|T_1 = c] \\
&= E[Y(1, D_2(1)) - Y(0, D_2(1))|T_1 = c] + E[Y(0, D_2(1)) - Y(0, D_2(0))|T_1 = c] = \theta_{c1}(1) + \delta_{c1}(0) \\
&= E[Y(1, D_2(0)) - Y(0, D_2(0))|T_1 = c] + E[Y(1, D_2(1)) - Y(1, D_2(0))|T_1 = c] = \theta_{c1}(0) + \delta_{c1}(1).
\end{aligned}$$

The notation $\theta_{c1}(1)$, $\theta_{c1}(0)$, $\delta_{c1}(1)$, $\delta_{c1}(0)$ makes explicit that direct and indirect effects may be heterogenous with respect to d_1, which permits interaction effects between D_1 and D_2 on Y. In the context of our health example, $\theta_{c1}(1)$ and $\theta_{c1}(0)$ are the direct effects of college attendance among first period compliers if their labor market states were set to their potential values with and without going to college.

Yamamoto (2013) shows identification of (4.2) and (4.3) based on an instrument for D_1 and a combined MAR and LI-type assumption, see the discussion in Section 4.1, with respect to D_2:

$$Y(d_1, d_2) \perp D_2(d_1')|Z, T_1 = c, X, \qquad \text{for } d_1, d_1' \in \{0, 1\}.$$

This allows controlling for the endogeneity of the latter despite the absence of a second instrument for D_2, given that conditional independence of D_2 holds given the compliance type and observed variables. While D_1 and its instrument are both assumed to be binary, D_2 might also be discretely multivalued or even continuous.

In contrast, Frölich and Huber (2014a) base identification on two distinct instruments Z_1 and Z_2 for D_1 and D_2, respectively. While Z_1 and D_1 are assumed to the binary, the authors consider various sets of assumptions that yield (4.2) and (4.3) under continuous Z_2, D_2, continuous Z_2 and discrete D_2, and discrete Z_2 and continuous D_2. Furthermore, they also discuss identification of the so-called controlled direct effect defined as the effect of D_1 when D_2 is fixed at a particular value d_2 for every complier (rather than its potential value $D_2(d_1)$), a parameter that also fits into dynamic treatment effects framework:

$$E[Y(1, d_2) - Y(0, d_2)|T_1 = c].$$

5 Violations and relaxations of the IV assumptions

As discussed in Section 2.1 in the context of the quarter of birth instrument, the standard IV assumptions or their conditional versions of Section 3 might be violated in many empirical contexts. Sections 5.1 and 5.2 discuss how violations of Assumptions 2.1 and 2.2 affect identification and under which relaxations causal effects on specific subpopulations can nevertheless be obtained. Troughout the section, we will assume that Assumption 2.3 holds.

5.1 Violation of the exclusion restriction

First, we analyze the Wald estimand under violations of the exclusion restriction $Y(1, d) = Y(0, d) = Y(d)$ for $d \in \{0, 1\}$ inherent in Assumption 2.1, while maintaining the independence (Assumption 2.1) and monotonicity (Assumption 2.2). First, consider a scenario under which there is no exclusion restriction for the noncompliers (i.e., always and never takers). Angrist

et al. (1996) show that the Wald estimand equals the LATE plus a bias term given by

$$\frac{E[Y(1,D(1))-Y(0,D(0)]}{E[D(1)-D(0)]}-E[Y(1,D(1))-Y(0,D(0))|T=c]=E[H|T\neq c]\cdot\frac{1-\pi_c}{\pi_c}$$

where $H=Y(1,d)-Y(0,d)$ denotes the causal effect of Z on Y. Second, let us consider a scenario in which there is not only a direct effect of Z on Y for noncompliers but also for compliers. In addition, suppose that $Y(1,0)-Y(0,0)=Y(1,1)-Y(0,1)$ for all compliers. The reason for imposing this additional "homogeneity" assumption is that it allows us to conventiently express the causal effect of Z on Y as H and the causal effect of D on Y as $G=Y(z,1)-Y(z,0)$. Using this additional notation, the IV estimand can be written as

$$\frac{E[Y(1,D(1))-Y(0,D(0)]}{E[D(1)-D(0)]}=E[G|T=c]+\frac{E[H]}{\pi_c},$$

see Angrist et al. (1996). The second term gives the bias relativ to the LATE and is

$$E[H|T=c]+E[H|T\neq c]\cdot\frac{1-\pi_c}{\pi_c}. \tag{5.1}$$

To interpret this result, let us consider the two components of the bias (5.1) separately. The first term originates from the direct effect of Z on Y for the compliers and does not depend on the compliance rate π_c. Thus, even under perfect compliance (that is if $\pi_c=1$) this part of the bias would prevail whereas the second part would be zero. The second term equals the product of the direct effect of Z on Y for noncompliers and the odds of being a noncomplier. This implies that the sensitivity of the IV estimand to violations of the exclusion restriction depends on the strength of the instrument as measured by the size of the compliant population. The exclusion restriction is therefore crucial for point identification. In Section 6.2, we present alternative approaches to obtain bounds on the LATE under violations of the exclusion restriction.

5.2 Violation and relaxations of monotonicity

When maintaining Assumption 2.1, but relaxing the monotonicity Assumption 2.2 such that defiers are permitted, the equality in (2.15) does not hold any more, see Angrist et al. (1996), but corresponds to

$$\frac{E(Y|Z=1)-E(Y|Z=0)}{E(D|Z=1)-E(D|Z=0)} = \Delta_c - \frac{\pi_d \cdot (\Delta_c - \Delta_d)}{\pi_c - \pi_d} = \frac{\pi_c \cdot \Delta_c - \pi_d \cdot \Delta_d}{\pi_c - \pi_d},$$

where $\Delta_d = E[Y(1) - Y(0)|T = d]$ denotes the LATE among defiers. Only in the special case that $\Delta_c = \Delta_d$ does this yield the LATE among compliers (and defiers). That is, if one is willing to replace the monotonicity assumption by homogeneity in average effects across complier and defier populations, causal effects are still identified.

When not imposing the strong and therefore typically not attractive effect homogeneity assumption, Δ_c is generally not identified under a violation of Assumption 2.2. This does, however, not necessarily mean that nothing can be said about the LATE at all. Small and Tan (2007) show that the sign of Δ_c is still identified if Assumption 2.2 is replaced by a somewhat weaker stochastic monotonicity condition, which is satisfied if $\Pr(T = c|Y(1), Y(0)) \geq P(T = d|Y(1), Y(0))$ (or $\Pr(T = c|U) \geq P(T = d|U)$ when assuming model (2.1)). That is, given any pair of potential outcome values under treatment and non-treatment, there must exist at least as many compliers as defiers. The same kind of assumption has also been considered in DiNardo and Lee (2011).

Several contributions even show that particular treatment effects can be point identified under specific relaxations of Assumption 2.2. Klein (2010) considers a nuisance term in the treatment equation that is unrelated with the potential outcomes and other unobserved factors affecting D (V in model (2.1)) and entails random departures from monotonicity such that some subjects defy. He discusses the conditions under which bias approximations for the identification of the LATE and MTE are obtained.

Secondly, de Chaisemartin (2016) shows that the Wald estimand identifies the LATE among a subpopulation of compliers, which he denotes as 'comvivors', if the following assumption is satisfied:

Assumption 5.1 (Compliers-defiers)**.** *There exists a subpopulation of compliers, denoted by* $T = cd$*, which satisfies* $\pi_{cd} = \pi_d$ *and* $E[Y(1) - Y(0)|T = cd] = \Delta_{cd} = \Delta_d$*.*

Assumption 5.1 states that some proportion of the total of compliers is equal to the defiers in terms of average effects and population size. Assumptions 2.1, 2.3 and 5.1 identify the LATE on the remaining compliers not necessarily resembling the defiers, the so-called comvivors, which are defined as $T = cv : c$ without cd, i.e. all compliers that outnumber those compliers resembling the defiers:

$$\frac{E(Y|Z=1) - E(Y|Z=0)}{E(D|Z=1) - E(D|Z=0)} = E[Y(1) - Y(0)|T = cv] \equiv \Delta_{cv}.$$

Therefore, the Wald estimator and TSLS still consistently estimate a causal effect as in the standard LATE framework, however, for a more local complier population.

Assumption 5.1 may appear abstract at the first glance, but de Chaisemartin (2016) discusses several restrictions that imply Assumption 5.1 and are easier to interpret. One possible condition is that compliers always outnumber defiers with the same treatment effect: $\Pr(T = c|Y(1) - Y(0)) \geq \Pr(T = d|Y(1) - Y(0))$, which is implied by, but weaker than the stochastic monotonicity assumption of Small and Tan (2007) discussed before. A second restriction is that the LATEs on defiers and compliers have the same sign and that the ratio of the LATEs is not 'too' large: Either $\text{sgn}\Delta_d = \text{sgn}\Delta_c \neq 0$ and $\Delta_d/\Delta_c \leq \pi_c/\pi_d$ or $\Delta_d = \Delta_c = 0$. de Chaisemartin (2016) gives several empirical examples in which Assumption 2.2 appears unrealistic but Assumption 5.1 is arguably likely satisfied, for instance in the evaluation of employment effects of disability insurance when average allowance rates of randomly assigned examiners serve as instrument as in Maestas et al. (2013). Similar arguments hold for studies of the effects of incarceration when using average sentencing rates of randomly assigned judges as an instrument for incarceration, see Aizer and Doyle (2013).

As a further strategy, Dahl et al. (2016) consider replacing Assumption 2.2 by a weaker local monotonicity condition given particular values of either marginal potential outcome distribution:

Assumption 5.2 (Local monotonicity)**.** *Either* $\Pr(T = d|Y(d) = y(d)) = 0$ *or* $\Pr(T = c|Y(d) = y(d)) = 0$ *for* $d \in \{0, 1\}$ *and all* $y(d)$ *in the support of* $Y(d)$*.*

While Assumption 5.2 allows for the presence of both compliers and defiers in the total population, it implies that conditional on a specific value of the potential outcome under treatment or non-treatment, either one or the other must not exist. Put differently, compliers and defiers are required to 'inhabit' different and non-overlapping regions of the marginal potential outcome distributions. Then, Δ_c is identified over all y satisfying $f(y, D = 1|Z = 1) - f(y, D = 1|Z = 0) > 0$ and $f(y, D = 0|Z = 0) - f(y, D = 0|Z = 1) > 0$ (see (2.12) and (2.13)), while Δ_d is based on all y for which $f(y, D = 1|Z = 1) - f(y, D = 1|Z = 0) < 0$ and $f(y, D = 0|Z = 0) - f(y, D = 0|Z = 1) < 0$. As an empirical example, reconsider the quarter of birth instrument for education and redshirting (postponement of school entry) as source of defiers, which more frequently occurs among families with a high socio-economic state, see Bedard and Dhuey (2006) and Aliprantis (2012). Assumption 5.2 would be satisfied if the socioeconomic status determined both defiance and the potential outcomes in a deterministic matter (e.g. children coming from defying families with a high socio-economic status can expect higher potential earnings than children of complying families).

In most applications including the quarter of birth instrument, a non-overlapping support in the potential outcomes of compliers and defiers appears unrealistic. However, when combining the ideas of local monotonicity and stochastic monotonicity, Assumption 5.2 can be weakened to an empirically more plausible local stochastic monotonicity condition:

Assumption 5.3 (Local stochastic monotonicity)**.** *For* $d \in \{1, 0\}$ *and* $y(d) \in supp(Y(d))$*:*

$\Pr(T = c|Y(d) = y(d)) \geq \Pr(T = d|Y(d) = y(d))$ *implies that*

$\Pr(T = c|Y(1) = y(1), Y(0) = y(0)) \geq \Pr(T = d|Y(1) = y(1), Y(0) = y(0))$*;*

and

$\Pr(T = c|Y(d) = y(d)) \leq \Pr(T = d|Y(d) = y(d))$ *implies that*

$\Pr(T = c|Y(1) = y(1), Y(0) = y(0)) \leq \Pr(T = d|Y(1) = y(1), Y(0) = y(0))$

This assumption allows for both compliers and defiers at any value of either marginal potential outcome distribution. However, it requires that if the share of one type weakly dominates the other conditional on either $Y(1)$ or $Y(0)$, it must also dominate conditional on both potential outcomes jointly, i.e. $Y(1)$ and $Y(0)$. de Chaisemartin (2012) demonstrates that under Assumptions 2.1, 2.3 and 5.3, the methods of Dahl et al. (2016) identify the LATEs on a subset

of compliers outnumbering the defiers whenever $\Pr(T = c|Y(1) = y(1), Y(0) = y(0)) \geq \Pr(T = d|Y(1) = y(1), Y(0) = y(0))$ and a subset of defiers outnumbering the compliers whenever $\Pr(T = c|Y(1) = y(1), Y(0) = y(0)) \leq \Pr(T = d|Y(1) = y(1), Y(0) = y(0))$.

6 Testing, sensitivity checks, and bounds

We subsequently discuss various approaches to test the IV assumptions, see Section 6.1, and outline sensitivity checks and bounds on the parameters of interest if one is not willing to maintain the satisfaction of Assumptions 2.1 – 2.3; see Section 6.2.

6.1 Testing the LATE assumptions

Under Assumptions 2.1 – 2.3, (2.12) and (2.13) not only permit evaluating local treatment effects, but also provide testable implications of the identifying assumptions. Namely, $f(y, D = 1|Z = 1) - f(y, D = 1|Z = 0) = f(y(1), T = c)$ and $f(y, D = 0|Z = 0) - f(y, D = 0|Z = 1) = f(y(0), T = c)$ imply for all y in the support of Y that

$$f(y, D = 1|Z = 1) \geq f(y, D = 1|Z = 0), \quad f(y, D = 0|Z = 0) \geq f(y, D = 0|Z = 1), \quad (6.1)$$

otherwise the joint densities of the compliers would be negative, even though a density has a lower bound of zero. Therefore, if one or both of the weak inequalities in (6.1) are violated, at least one of Assumptions 2.1 – 2.3 is violated. These constraints were first derived by Balke and Pearl (1997) for binary outcomes, while Heckman and Vytlacil (2005) formulated them in terms of continuous outcomes. Note that the testable implications (6.1) remain unchanged when easing Assumption 2.2 to stochastic monotonicity of the form $\Pr(T = c|Y(d)) \geq \Pr(T = d|Y(d))$ for $d \in \{1, 0\}$, see Mourifié and Wan (2014).

For testing, (6.1) could be verified at each value y in the support of Y. However, if the outcome is of rich support (e.g., continuous), finite sample power may be higher when partitioning

the support into a finite number of subsets. The testable constraints then are

$$\begin{aligned}
&\Pr(Y \in A, D = 1|Z = 1) \geq \Pr(Y \in A, D = 1|Z = 0), \\
&\Pr(Y \in A, D = 0|Z = 0) \geq \Pr(Y \in A, D = 0|Z = 1),
\end{aligned} \tag{6.2}$$

where A denotes a subset of the support of Y. Kitagawa (2015) proposes a test based on re-sampling a variance-weighted two sample Kolmogorov-Smirnov-type statistic on the supremum of $\Pr(Y \in A, D = 1|Z = 0) - \Pr(Y \in A, D = 1|Z = 1)$ and $\Pr(Y \in A, D = 0|Z = 1) - \Pr(Y \in A, D = 0|Z = 0)$, respectively, across multiple subsets A. The method can also be used for testing conditional on observed covariates, if the latter are binned into subsets of the support in a similar way as the outcomes. As an alternative approach, Mourifié and Wan (2014) show that a modified version of (6.2) making use of conditional moment inequality constraints fits the intersection bounds framework of Chernozhukov et al. (2013b). For this reason, the corresponding STATA package of Chernozhukov et al. (2013a) can for instance be used to implement the test either unconditionally or conditional on observed covariates. For binary outcomes, Machado et al. (2013) propose a procedure that both verifies the constraints and the sign of the average treatment effect on the entire population.

As an alternative set of testable constraints, Huber and Mellace (2015) show that the LATE assumptions imply the following restrictions related to the mean potential outcomes (i) of the always takers under treatment and (ii) of the never takers under non-treatment:

$$\begin{aligned}
&E(Y|D = 1, Z = 1, Y \leq y_q) \leq E(Y|D = 1, Z = 0) \leq E(Y|D = 1, Z = 1, Y \geq y_{1-q}), \\
&E(Y|D = 0, Z = 0, Y \leq y_r) \leq E(Y|D = 0, Z = 1) \leq E(Y|D = 0, Z = 0, Y \geq y_{1-r}).
\end{aligned} \tag{6.3}$$

Under Assumptions 2.1 – 2.2, $q = \Pr(D = 1|Z = 0)/\Pr(D = 1|Z = 1)$ gives the share of always takers among those with $D = 1$ and $Z = 1$, i.e., in the mixed population of compliers and always takers, and y_q is the qth quantile of Y given $D = 1$ and $Z = 1$. $r = 1 - (\Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0))/\Pr(D = 0|Z = 0)$ corresponds to the share of never takers among those with $D = 0$ and $Z = 0$, and y_r is the r^{th} quantile of Y given $D = 0$ and $Z = 0$. Considering the

first line of (6.3), the intuition of the test is as follows: $E(Y|D=1,Z=0)$ point identifies the mean potential outcome of the always takers under treatment, as any subject with $D=1$, $Z=0$ must be an always taker in the absence of defiers. Furthermore, the mean potential outcomes of the always takers are bounded by the averages in the upper and lower outcome proportions with $D=1$ and $Z=1$ that correspond to the share of the always takers in the mixed population: $E(Y|D=1,Z=1,Y\leq y_q)$, $E(Y|D=1,Z=1,Y\geq y_{1-q})$. $E(Y|D=1,Z=0)$ must lie within the latter bounds, otherwise the identifying assumptions are necessarily violated. An analogous result applies to the mean potential outcome of never takers under non-treatment. Any procedure suitable for testing multiple moment inequalities could be used for verifying (6.3), for instance the method by Chen and Szroeter (2014).

While it appears attractive to have tests of the IV assumptions even in the just identified case with one instrument and one treatment, it needs to be pointed out that any of the tests discussed so far check for necessary, albeit not sufficient conditions. That is, the tests are inconsistent in the sense there may exist counterfactual distributions which satisfy the testable restrictions, but violate Assumptions 2.1 – 2.3.[1] Sharma (2016) offers an extension to merely testing (6.2) by determining the likelihood an instrument is valid when the testable constraints are satisfied. To this end, the test defines classes of valid causal models satisfying Assumptions 2.1 – 2.3 as well as as invalid models and compares their marginal likelihood in the observed data.

Several further tests that are not based on constraints (6.2) have been proposed. Slichter (2014) suggests testing conditional IV independence (Assumption 3.1) by finding covariate values $X=x$ for which Z has no first stage and checking whether Z is associated with Y despite the absence of a first stage. For the multivalued treatment case as discussed in Section 2.4, Angrist and Imbens (1995) argue that Assumption 2.2 implies that the cdfs of D given $Z=1$ and $Z=0$, respectively, do not cross (i.e., stochastic dominance),

$$\Pr(D\geq j|Z=1)\geq\Pr(D\geq j|Z=0),\quad D,j\in\{0,1,...,J\}, \tag{6.4}$$

[1]Interestingly, asymptotic power ceteris paribus increases as the share of compliers decreases, i.e. as the instrument becomes weaker. Therefore, the tests supposedly have very low power in randomized trials with a large first stage, where, however, the LATE assumptions often appear quite credible. On the other hand, the tests might be used in quasi-experimental settings where the assumptions are more challengeable and the first stage is small, as in the case of the quarter of birth instrument.

which may be verified in the data. Fiorini and Stevens (2014) point out that testing the necessary (albeit not sufficient) condition (6.4) can have power against violations of both Assumption 2.2 and the independence of Z and $D(1), D(0)$ which is part of Assumption 2.1. In the presence of both a binary and a continuous instrument, Dzemski and Sarnetzki (2014) suggest a nonparametric overidentification tests for IV validity. Finally, if outcome variables are observed already in periods prior to instrument and treatment assignment, placebo tests based on estimating the effect of Z on pre-instrument outcomes may be performed to check the plausibility of Assumption 2.1.

6.2 Sensitivity checks and bounds

If IV validity appears dubious or has even been refuted by the tests presented in Section 6.1, one may consider sensitivity checks on the robustness of the LATE under violations of the IV assumptions or the derivation of upper and lower bounds on the effect (rather than point identification) under weaker assumptions. Huber (2014), for instance, proposes sensitivity checks under the non-satisfaction of the IV exclusion restriction (inherent in Assumption 2.1) or Assumption 2.2. Under a presumed violation of the exclusion restriction while maintaining the random assignment of Z and Assumptions 2.2 and 2.3, the LATE can be shown to correspond to

$$\frac{E(Y|Z=1) - E(Y|Z=0) - \Pr(D=1|Z=0)\cdot\gamma_a - \Pr(D=0|Z=1)\cdot\gamma_n}{\Pr(D=1|Z=1) - \Pr(D=1|Z=0)} - \gamma_c, \tag{6.5}$$

where γ_c, γ_a, and γ_n denote the potentially heterogenous average direct effects of Z to the mean potential outcomes of the compliers, always takers, and never takers, respectively. Jones (2015) derives a related result under the assumption that the direct effect on the never takers (γ_n) is equal to zero. Under a homogeneous direct mean effect across types, implying that $\gamma_a = \gamma_n = \gamma_c = \gamma$, (6.5) simplifies to $(E(Y|Z=1) - E(Y|Z=0) - \gamma)/(\Pr(D=1|Z=1) - \Pr(D=1|Z=0))$. Inference can therefore be conducted if the researcher has a plausible prior about the possible range of values γ_c, γ_a, γ_n, or γ might take, see also Conley et al. (2012). The approach suggested by Slichter (2014) based on his IV validity test (see Section 6.1) may be used for determining such values in sensitivity checks.

Under a violation of Assumption 2.2 while maintaining Assumptions 2.1 and 2.3, Huber (2014) shows that the mean potential outcomes of the compliers correspond to the following expressions:

$$\begin{aligned} E(Y(1)|T=c) &= \frac{\Pr(D=1|Z=1)\cdot E(Y|D=1,Z=1)}{\rho_a(\Pr(D=1|Z=0)-\pi_d)+\Pr(D=1|Z=1)-\Pr(D=1|Z=0)+\pi_d}, \\ E(Y(0)|T=c) &= \frac{\Pr(D=0|Z=0)\cdot E(Y|D=0,Z=0)}{\rho_n(\Pr(D=0|Z=1)-\pi_d)+\Pr(D=1|Z=1)-\Pr(D=1|Z=0)+\pi_d}, \end{aligned} \tag{6.6}$$

where $\rho_a = E(Y(1)|T=a)/E(Y(1)|T=c)$ and $\rho_n = E(Y(0)|T=n)/E(Y(0)|T=c)$ are the ratios of mean potential outcomes (i) of always takers and compliers under treatment and (ii) of never takers and compliers under non-treatment, respectively. Considering various combinations of ρ_a, ρ_n, π_d allows investigating the sensitivity of the LATE on compliers to violations of monotonicity.

If plausible values for the aforementioned tuning parameters appear hard to justify, a bounds analysis may appear more credible, at the potential cost of getting a larger set of potential values for the LATE. Flores and Flores-Lagunes (2013) maintain Assumption 2.2 and the random assignment of Z, but assume the violation of the exclusion restriction. They instead impose restrictions on the order of specific mean potential outcomes (i) across treatment or instrument states within particular types and (ii) across types to narrow the bounds. Mealli and Pacini (2013) show that bounds can alternatively be tightened if an auxiliary variable is at hand for which the exclusion restriction (contrary to the outcome of interest) holds, as for instance a covariate measured prior to randomization, and which is associated with the outcome and/or the compliance type. Richardson and Robins (2010) maintain Assumption 2.1, but assume a violation of Assumption 2.2 and derive bounds for the LATEs of various compliance types when the outcome is binary. Under mean independence of Z and the potential outcomes/treatment states, Huber et al. (2014) bound the LATEs on several subpopulations (also for non-binary outcomes) when monotonicity is violated, with and without invoking a particular order in the mean potential outcomes across types.

7 External validity of the LATE

Whether the LATE or other local effects are relevant parameters given that they only refer to the subpopulation of compliers heavily depends on the empirical context and has been controversially discussed in the literature, see for instance Imbens (2010b), Deaton (2010), and Heckman and Urzúa (2010). In treatment evaluation, one typically strives for the identification of 'global' effects on the entire or the treated population. One could therefore argue that the relevance of the LATE crucially depends on its external validity, i.e. its similarity to the a priori unidentified average treatment effect (ATE) in the entire population. For this reason, we subsequently discuss potential checks for external validity based on observables (Section 7.1), conditions for extrapolating the LATE to the ATE and tests thereof (Section 7.2), and partial identification of the ATE based on the IV assumptions and further restrictions (Section 7.3).

7.1 Comparability in terms of observables

Comparing compliers and the total population in terms of observed characteristics may be useful for judging the plausibility of the LATE being (close to) externally valid. Angrist and Fernández-Val (2010) consider $\Pr(X = x|T = c)/\Pr(X = x)$, the relative likelihood of covariate values $X = x$ among compliers compared to the entire population, which is identified under Assumptions 2.1 – 2.3 by the ratio of the first stage given $X = x$ to the overall first stage:

$$\frac{\Pr(X = x|T = c)}{\Pr(X = x)} = \frac{\Pr(T = c|X = x)}{\Pr(T = c)} = \frac{E(D|Z = 1, X = x) - E(D|Z = 0, X = x)}{E(D|Z = 1) - E(D|Z = 0)}.$$

Furthermore, under the conditional IV assumptions (Assumptions 3.1 – 3.4), the mean or other distributional features of the covariates among compliers can be obtained by using the κ-weighting function of Abadie (2003) provided in Section 3.5:

$$E(X|T = c) = \frac{E(\kappa \cdot X)}{E(\kappa)}.$$

While such checks may provide important insights about the representativeness of compliers in terms of X, the important caveat remains that nothing can be said about unobserved charac-

teristics which may be related to the potential outcomes.

7.2 Conditions for extrapolation and testing

This section discusses various structural assumptions that allow extrapolating from the LATE to the ATE. Angrist (2004) distinguishes two restrictions in which the LATE is directly externally valid, i.e. corresponds to the ATE. Under the first restriction, there is no selection in the sense that mean potential outcomes under either treatment state are constant across types. Under the second restriction, selection is of a rather specific form such that the levels of the mean potential outcomes differ across types, but the mean effects do not, i.e. are homogeneous. Note that under the first restriction, both (i) the Wald estimand and (ii) $E(Y|D=1)-E(Y|D=0) = E(Y(1)|T \in \{c,a\})-E(Y(0)|T \in \{c,n\})$ identify the ATE because $E(Y(d)|T) = E(Y(d))$, while under the second restriction, only (i) but not (ii) yields the effect, as $E(Y(d)|T) \neq E(Y(d))$. Angrist (2004), Brinch et al. (2012), and Huber (2013) consider tests for the external validity of the LATE under the first restriction based on differences in mean potential outcomes across types. To see the intuition, consider the following regression representation, which is fully nonparametric as D, Z are binary:

$$E(Y|D,Z) = \beta_0 + \beta_D D + \beta_Z Z + \beta_{DZ} DZ,$$

with

$$\begin{aligned}
\beta_Z &= E(Y|D=0,Z=1) - E(Y|D=0,Z=0) = E(Y(0)|T=n) - E(Y(0)|T \in \{c,n\}), \\
\beta_{DZ} &= E(Y|D=1,Z=1) - E(Y|D=0,Z=1) - \{E(Y|D=1,Z=0) - E(Y|D=0,Z=0)\} \\
&= E(Y(1)|T \in \{c,a\}) - E(Y(0)|T=n) - \{E(Y(1)|T=a) - E(Y(0)|T \in \{c,n\})\}.
\end{aligned}$$

β_Z captures selection driven by the difference in mean potential outcomes of compliers and never takers under non-treatment (or differences in U under model (2.1)). β_{DZ} reflects both selection and treatment effect heterogeneity across types.

If $\beta_Z = 0$, β_{DZ} simplifies to $E(Y(1)|T \in \{c,a\}) - E(Y(1)|T=a)$, and a difference may be

due to selection or differential treatment effects across always takers and compliers. Therefore, jointly testing for β_Z and β_{DZ} has in general non-trivial power to detect heterogeneity in mean potential outcomes across types, either driven by selection or treatment effect heterogeneity, which generally implies that the LATEs differ across types, too. However, not all potential violations can be tested, as the potential outcome of never takers under treatment and always takers under non-treatment is never observed. Strictly speaking, $\beta_Z = \beta_{DZ} = 0$ is therefore not sufficient for external validity under the first restriction of Angrist (2004). Furthermore, it is not necessary either, because the LATE can theoretically be homogenous across groups even if potential outcomes differ (second restriction of Angrist (2004)). However, under the assumption that differences in mean potential outcomes either occur across all or across no types (i.e. either $E(Y(d)|T) = E(Y(d))$ or $E(Y(d)|T = t) \neq E(Y(d)|T = t')$ for any $t \neq t'$ and $t, t' \in \{c, a, n\}$), β_Z suffices for detecting selection. Conditional on $\beta_Z = 0$, β_{DZ} in this case exclusively detects effect heterogeneity across types. For this reason, it appears worthwhile testing $\beta_Z = \beta_{DZ} = 0$, which can be easily implemented by means of an F-test.[2] In the case of one-sided noncompliance ($\Pr(D(0) = 1) = 0$), only $E(Y|D = 0, Z = 1) - E(Y|D = 0, Z = 0) = 0$ is testable.

If Assumptions 3.1 – 3.4, rather than Assumptions 2.1 – 2.3 are satisfied, the same type of test may be conducted conditional on X, see de Luna and Johansson (2014) and Black et al. (2015). In this case, testing can also be framed as a check for the conditional mean independence of the treatment and the potential outcomes given observed covariates: $E(Y(d)|D, X) = E(Y(d)|X)$, which would imply that $E(Y|D = 1, X) - E(Y|D = 0, X)$ yielded the causal effect $E(Y(1) - Y(0)|X)$. Donald et al. (2014b) suggest an alternative approach to test this condition based on a comparison of treatment effects under one-sided noncompliance, ruling out always takers and defiers. As in the latter case the LATT ($\Delta_{c,D=1}$) corresponds to the ATT ($\Delta_{D=1}$), Donald et al. (2014b) construct a Durbin-Wu-Hausmann-type test based on the z-statistic for the difference of the respective estimates $\hat{\Delta}_{c,D=1}$ and $\hat{\Delta}_{D=1}$. These estimates are obtained by the sample analogue of (3.6) and the approach suggested in Hirano et al. (2003), respectively. While the satisfaction of conditional mean independence of the treatment implies

[2]The following approaches are asymptotically equivalent to this F-test: Testing (i) $\frac{E(Y|Z=1)-E(Y|Z=0)}{E(D|Z=1)-E(D|Z=0)} = E(Y|D = 1) - E(Y|D = 0)$ as in the classical Hausman (1978) test, (ii) $\frac{E(Y|Z=1)-E(Y|Z=0)}{E(D|Z=1)-E(D|Z=0)} = E(Y|D = 1, Z = 0) - E(Y|D = 0, Z = 1)$ as suggested in Angrist (2004), and (iii) $E(Y|D = 1, Z = 1) = E(Y|D = 1, Z = 0)$, $E(Y|D = 0, Z = 1) = E(Y|D = 0, Z = 0)$ as considered in the context of fuzzy regression discontinuity designs by Bertanha and Imbens (2015).

that identification and extrapolation can be achieved without an instrument, the availability of an instrument is required to obtain an overidentifying restriction and being able to construct a test for $E(Y(d)|D,X) = E(Y(d)|X)$. Note that if the latter assumption holds, not only the LATE on compliers, ATE, and ATT are identified, but also the LATEs on never and always takers, as discussed in Frölich and Lechner (2015).

If the strong assumption of effect homogeneity across types is not satisfied, the LATE may nevertheless permit extrapolation to the ATE even under a binary instrument if particular parametric assumptions hold. This is shown in Brinch et al. (2012), who assume the MTE $\Delta(\bar{V} = p(z))$ of Section 2.3 to be linear in $p(Z) = \Pr(D = 1|Z)$ by imposing linearity on $E(Y(0)|p(Z))$ and $E(Y(1)|p(Z))$, see also Restriction 3 in Angrist (2004). Brinch et al. (2012) further demonstrate that polynomial (rather than linear) MTE functions are identified if the MTE is additively separable in X and unobservable factors and if X satisfies specific support conditions.

As an alternative source of extrapolation, Angrist and Fernández-Val (2010) and Aronow and Carnegie (2013) consider homogeneity of the average treatment effects given X across types, a conditional version of (unconditional) effect homogeneity under the second restriction of Angrist (2004):

Assumption 7.1 (Conditional effect homogeneity)**.** $E(Y(1)-Y(0)|T,X) = E(Y(1)-Y(0)|X) = \Delta(x)$

Assumption 7.1 implies that heterogeneity in average effects across types is solely due to X, such that $\Delta_c(x) = \Delta(x)$, implying that the ATE, denoted as Δ, is obtained by $\int \Delta_c(x) dF_X(x) = E(\Delta_c(x))$ if the conditional LATE assumptions (Assumptions 3.1 – 3.4) are satisfied.[3] More generally, the ATE on some population selected by the binary indicator S (e.g. $S = D$ for the treated and $S = 1 - D$ for the nontreated) corresponds to

$$\Delta_{S=1} = \int \Delta_c(x) dF_{X|S=1}(x) = \int \Delta_c(x) \frac{\Pr(S=1|X)}{\Pr(S=1)} dF_X(x) = E\left[\Delta_c(x) \frac{\Pr(S=1|X)}{\Pr(S=1)}\right].$$

It is important to note that conditional effect homogeneity rules out that effect heterogene-

[3] In fact, under Assumption 7.1, Assumption 3.2 might even be relaxed for instance to stochastic monotonicity given X.

ity is driven by unobserved gains, which importantly restricts the source of treatment effect heterogeneity and is, for instance, not consistent with the Roy (1951) model.

Angrist and Fernández-Val (2010) demonstrate that Assumption 7.1 is testable (conditional on the satisfaction of Assumptions 3.1 – 3.4) if more than one instrument is available. Denote by $\Delta_c^W(x)$ and $\Delta_c^Z(x)$ the conditional LATEs based on two different instruments W and Z. It must hold that

$$\Delta_{S=1} = \int \Delta_c^W(x) \frac{\Pr(S=1|X)}{\Pr(S=1)} dF_X(x) = \int \Delta_c^Z(x) \frac{\Pr(S=1|X)}{\Pr(S=1)} dF_X(x).$$

See also Heckman et al. (2010) for testing approaches in the context of the MTE framework that verify conditional effect homogeneity based on multiple instruments.

Another extrapolation strategy is based on the rank invariance assumption discussed in Section 8.2. Vuong and Xu (2014) and Wüthrich (2016) show that under rank invariance, the counterfactual mappings,

$$P_{01|T=t} \equiv Q_{Y(0)|T=t}\left(F_{Y(1)|T=t}(y)\right) \quad \text{and} \quad P_{10|T=t} \equiv Q_{Y(1)|T=t}\left(F_{Y(0)|T=t}(y)\right),$$

which relate each individual outcome to its counterfactual, do not depend on the type $T = t$. Hence, one can use $P_{01|T=c}(y)$ and $P_{10|T=c}(y)$, which are both identified under Assumptions 2.1 – 2.3, for imputing the counterfactual distributions of $Y(1)$ for never takers and of $Y(0)$ for the always-takers. This is exactly the intuition underlying the instrumental variable quantile regression model discussed in Section 8.2.

7.3 Partial identification of the ATE

Even if point identification of the ATE fails because the LATE estimates are not externally valid, the identifying power of the IV assumptions may still be used to at least partially identify the ATE and other parameters such as the ATT not discussed here. Balke and Pearl (1997) (for binary outcomes) as well as Heckman and Vytlacil (2001a) and Kitagawa (2009) (for more general outcomes) derive bounds on the ATE under Assumptions 2.1 – 2.3 and also provide the interesting result that they coincide with the bounds of Manski (1990), who merely assumes

$E(Y(d)|Z = 1) = E(Y(d)|Z = 0)$ for $d \in \{1, 0\}$. Shaikh and Vytlacil (2011) sharpen the bounds on the ATE in the binary outcome case under the assumption that the treatment effect is either weakly positive or weakly negative for all individuals (while the direction is a priori not restricted). Cheng and Small (2006) extend the results for binary outcomes to the case that the treatment can take three values under particular forms of (one-sided) noncompliance.

Under mean independence of Z and the potential outcomes/treatment states and Assumption 2.2, Huber et al. (2014) bound the ATE when assuming a particular order in the mean potential outcomes across types. Also Flores et al. (2015) consider such restrictions in addition to Assumptions 1 and 2, but also invoke a specific order of mean potential outcomes across treatment states within specific types. Furthermore, see Chiburis (2010) and references therein for the derivation of semiparametric (rather than nonparametric) bounds on the ATE under the IV assumptions. Kowalski (2016) considers the MTE framework and assumes the marginal outcomes under treatment and non-treatment, $E(Y(1)|V = v)$ and $E(Y(0)|V = v)$, to be monotonic in the unobserved term in the treatment model (2.1) to bound the ATE. Angrist (2004) offers a sensitivity check for the ATE based on particular proportionality conditions across the mean potential outcomes of various types. Finally, Mogstad et al. (2016) develop a framework for obtaining identified sets on the ATE and other policy relevant parameters by exploiting the fact that the IV estimand and many other parameters of interest can be expressed as a weighted average of MTE where the weights are known or identified.

8 Relationship to other instrumental variable approaches

In this section, we discuss the relationship between the LATE framework and two other widely used IV models: the classical linear IV model and the instrumental variable quantile regression model (IVQR) due to Chernozhukov and Hansen (2005).

8.1 Linear IV models

Linear IV models such as

$$Y = X'\gamma + \beta D + \varepsilon$$

play a central role in applied empirical research. If we are willing to assume that treatment effects are homogeneous across individuals, the coefficient β corresponds to the population ATE, which can be consistently estimated using classical estimators such as TSLS or limited information maximum likelihood (LIML). However, in most applications it appears implausible that treatment effects are homogeneous and thus unrelated to observable or unobservable characteristics. It is therefore important to understand which parameters classical estimators of the linear IV model such as TSLS and LIML estimate when treatment effects are in fact heterogeneous.

To formalize the analysis, we follow Angrist and Imbens (1995) and Angrist and Pischke (2009) and consider the TSLS estimand with fully saturated first and second stage equations

$$D = \pi_X + \pi_{1X} Z + u, \quad Y = \alpha_X + \beta D + \varepsilon,$$

where π_X and α_X denote saturated models for covariates and π_{1X} denotes separate first-stage effects of Z for every value of X. Under the assumptions of the LATE framework with covariates (Assumption 3.1 – Assumption 3.4) it can be shown that

$$\beta = E\left(E\left(Y(1) - Y(0)|T = c, X\right)\omega(X)\right),$$

where

$$\omega(X) = \frac{Var\left(E(D|X,Z)|X\right)}{E\left(Var\left(E(D|X,Z)|X\right)\right)}.$$

That is, TSLS with a fully staturated first stage and a second stage which is saturated in the covariates produces a weighted average of covariate-specific LATEs with weights proportional to the average conditional variance of the population first-stage fitted value $E(D|X,Z)$.

Kolesar (2013) generalized the analysis in Angrist and Imbens (1995) by characterizing the estimands of general two-step estimators (such as TSLS) and minimum distance estimators (such as LIML) under the LATE framework. His analysis shows that while the probability limit of TSLS can be expressed as a convex combination of LATEs as in the special case discussed before, LIML and related estimators may end up being outside the convex hull of LATEs. As

a consequence, minimum distance estimands may not correspond to a causal effect if treatment effects are heterogenous.

8.2 IV quantile regression

The instrumental variable quantile regression (IVQR) model introduced by Chernozhukov and Hansen (2005) provides an alternative framework for identifying and estimating heterogeneous treatment effects with IVs. In contrast to the LATE framework, the IVQR model does not impose a monotonicity assumption in the selection equation. Instead, it relies on rank invariance in the outcome equation, a restriction on the evolution of individual ranks across treatment states.[4] By virtue of the rank invariance assumption, the IVQR model identifies population level treatment effects. This is in sharp contrast to the LATE framework under which treatment effects are only identified for the compliers. However, rank invariance substantially restricts treatment effect heterogeneity and may therefore be implausible in many applications. For instance, as noted by Heckman and Vytlacil (2007), rank invariance rules out scenarios in which agents self-select based on their individual effects and does not allow for effect heterogeneity as generated by the generalized Roy model.

To formalize the rank invariance assumption, note that by the Skorohod representation of random variables, the potential outcome $Y(d)$ can be related to its quantile function $Q_{Y(d)}(\tau)$ in the following way:

$$Y(d) = Q_{Y(d)}(U(d)), \quad \text{where} \quad U(d) \sim \text{Uniform}(0,1). \tag{8.1}$$

If the potential outcomes are continuous random variables, $Q_{Y(d)}(\cdot)$ is strictly increasing and the disturbance $U(d)$ determines the individual position or rank in the distribution of $Y(d)$. We therefore refer to $U(d)$ as 'rank'. With this notation at hand, we can formally define rank invariance as $U(1) = U(0)$. Under rank invariance and instrument independence, the population QTE, $\Delta(\tau) = Q_{Y(1)}(\tau) - Q_{Y(0)}(\tau)$, can be identified and estimated based on the

[4] Chernozhukov and Hansen (2005) show that rank invariance can be somewhat relaxed to rank similarity that allows for random deviations from the expected rank.

following conditional moment restriction:

$$\Pr\left(Y \leq Q_{Y(D)}(\tau)|Z\right) = \tau \tag{8.2}$$

On the surface, the IVQR and the LQTE model do not seem to be connected since they rely on different non-nested assumptions and identify treatment effects for different populations. Despite these differences, Wüthrich (2016) shows that

$$\Delta(\tau) = \Delta_c\left(F_{Y(0)|T=c}\left(Q_{Y(0)}(\tau)\right)\right) = \Delta_c\left(F_{Y(1)|T=c}\left(Q_{Y(1)}(\tau)\right)\right), \tag{8.3}$$

where $Q_{Y(1)}(\tau)$ and $Q_{Y(1)}(\tau)$ are defined as

$$\begin{aligned} Q^{-1}_{Y(1)}(y) &= \pi_a F_{Y(1)|T=a}(y) + \pi_c F_{Y(1)|T=c}(y) + \pi_n F_{Y(0)|T=n}\left(Q_{Y(0)|T=c}\left(F_{Y(1)|T=c}(y)\right)\right) \\ Q^{-1}_{Y(0)}(y) &= \pi_n F_{Y(0)|T=n}(y) + \pi_c F_{Y(0)|T=c}(y) + \pi_a F_{Y(1)|T=a}\left(Q_{Y(1)|T=c}\left(F_{Y(0)|T=c}(y)\right)\right). \end{aligned}$$

Equation (8.3) demonstrates that the IVQR QTE estimand at quantile level τ corresponds to the LQTE at τ', where τ will generally not be equal to τ'. The difference between the two estimates is determined by two factors: (i) the differences between the potential outcome distributions of the untreated compliers and never takers as well as the differences between the potential outcome distributions of the treated compliers and always takers, and (ii) the relative size of the three subpopulations.

The results in Wüthrich (2016) confirm that with unrestricted treatment effect heterogeneity, all the information on treatment effects has to come from the compliers, i.e. the group observed under either treatment state. Moreover, they provide insights on how the IVQR model extrapolates from the compliers to the whole population. This motivates the use of the IVQR as an approach to extrapolation in the LQTE framework; see Section 7.2.

9 Conclusion

This paper provides a survey on the methodological advancements in the evaluation of local average treatment effects based on instruments. We first review the classical framework going back to the seminal contributions of Imbens and Angrist (1994) and Angrist et al. (1996), which have been very influential in applied empirical research. We then proceed by summarizing and synthesizing important methodological extensions, for example distributional and quantile treatment effects, multivalued or multiple treatments and instruments, identification and estimation in the presence of observed covariates, attrition and measurement error, testing and relaxations of identifying assumptions, conditions for external validity, and the relationship to other IV approaches. We thereby complement more introductory reviews that focus on implementation and applications such as Imbens (2014) and the textbook discussions in Angrist and Pischke (2009) and Angrist and Pischke (2015).

References

Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association 97, 284–292.

Abadie, A., 2003. Semiparametric instrumental variable estimation of treatment response models. Journal of Econometrics 113, 231–263.

Abadie, A., Angrist, J., Imbens, G. W., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. Econometrica 70, 91–117.

Aizer, A., Doyle, J. J., 2013. Juvenile incarceration, human capital and future crime: Evidence from randomly-assigned judges. Technical report, NBER.

Aliprantis, D., 2012. Redshirting, compulsory schooling laws, and educational attainment. Journal of Educational and Behavioral Statistics 37, 316–338.

Angrist, J., Fernández-Val, I., 2010. Extrapolate-ing: External validity and overidentification in the late framework. NBER working paper 16566.

Angrist, J., Imbens, G. W., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of American Statistical Association 90, 431–442.

Angrist, J., Imbens, G. W., Rubin, D., 1996. Identification of causal effects using instrumental variables. Journal of American Statistical Association 91, 444–472 (with discussion).

Angrist, J., Krueger, A., 1991. Does compulsory school attendance affect schooling and earnings? Quarterly Journal of Economics 106, 979–1014.

Angrist, J. D., 2004. Treatment effect heterogeneity in theory and practice. The Economic Journal 114, C52–C83.

Angrist, J. D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Epiricist's Companion. Princeton University Press.

Angrist, J. D., Pischke, J.-S., 2015. Mastering 'Metrics: The Path from Cause to Effect. Princeton University Press.

Aronow, P. M., Carnegie, A., 2013. Beyond late: Estimation of the average treatment effect with an instrumental variable. Political Analysis 21, 492–506.

Balke, A., Pearl, J., 1997. Bounds on treatment effects from studies with imperfect compliance. Journal of the American Statistical Association 92, 1171–1176.

Barua, R., Lang, K., 2009. School entry, educational attainment, and quarter of birth: A cautionary tale of late. NBER Working Paper 15236.

Bedard, K., Dhuey, E., 2006. The persistence of early childhood maturity: International evidence of long-run age effects. The Quarterly Journal of Economics 121, 1437–1472.

Behaghel, L., Crépon, B., Gurgand, M., 2013. Robustness of the encouragement design in a two-treatment randomized control trial. IZA Discussion Paper No 7447.

Belloni, A., Chernozhukov, V., Fernández-Vál, I., Hansen, C., 2014. Program evaluation with high-dimensional data. cemmap Working Papers, 33/13.

Bertanha, M., Imbens, G., 2015. External validity in fuzzy regression discontinuity designs. NBER working paper 20773.

Black, D. A., Joo, J., LaLonde, R. J., Smith, J. A., Taylor, E. J., 2015. Simple tests for selection bias: Learning more from instrumental variables. IZA Discussion Paper No 9346.

Blackwell, M., 2015. Identification and estimation of joint treatment effects with instrumental variables. working paper, Department of Government, Harvard University.

Bloom, H. S., 1984. Accounting for no-shows in experimental evaluation designs. Evaluation Review 8, 225–246.

Bound, J., Jaeger, D. A., Baker, R. M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Brinch, C. N., Mogstad, M., Wiswall, M. J., 2012. Beyond late with a discrete instrument. heterogeneity in the quantity-quality interaction of children. forthcoming in the Journal of Political Economy.

Buckles, K. S., Hungerman, D. M., 2013. Season of birth and later outcomes: Old questions, new answers. Review of Economics and Statistics 95, 711–724.

Card, D., 1995. Using geographic variation in college proximity to estimate the return to schooling. In: Christofides, L., Grant, E., Swidinsky, R. (Eds.), Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp. University of Toronto Press, Toronto, pp. 201–222.

Carneiro, P., Heckman, J. J., Vytlacil, E. J., 2011. Estimating marginal returns to education. American Economic Review 101, 2754–2781.

Carneiro, P., Lee, S., 2009. Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality. Journal of Econometrics 149 (2), 191–208.

Chalak, K., 2016. Instrumental variables methods with heterogeneity and mismeasured instruments. Econometric Theory, 1–36.

Chalak, K., White, H., 2011. An extended class of instrumental variables for the estimation of causal effects. Canadian Journal of Economics 44, 1–51.

Chen, L.-Y., Szroeter, J., 2014. Testing multiple inequality hypotheses: A smoothed indicator approach. Journal of Econometrics 178, 678–693.

Chen, X., Flores, C. A., 2015. Bounds on treatment effects in the presence of sample selection and noncompliance: the wage effects of job corps. Journal of Business & Economic Statistics 33, 523–540.

Cheng, J., Small, D. S., 2006. Bounds on causal effects in three-arm trials with non-compliance. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (815-836).

Chernozhukov, V., Hansen, C., 2005. An iv model of quantile treatment effects. Econometrica 73, 245–261.

Chernozhukov, V., Kim, W., Lee, S., Rosen, A., 2013a. Implementing intersection bounds in stata. Centre for Microdata Methods and Practice Working Paper CWP38/13, Institute for Fiscal Studies.

Chernozhukov, V., Lee, S., Rosen, A., 2013b. Intersection bounds: Estimation and inference. Econometrica 81, 667–737.

Chiburis, R. C., 2010. Semiparametric bounds on treatment effects. Journal of Econometrics 159, 267ï¿½275.

Conley, T. G., Hansen, C. B., Rossi, P. E., 2012. Plausibly exogenous. Review of Economics and Statistics 94, 260–272.

Cornelissen, T., Dustmann, C., Raute, A., Uta, S., 2016. From late to MTE: Alternative methods for the evaluation of policy interventions. IZA DP No. 10056.

Dahl, C. M., Huber, M., Mellace, G., 2016. It's never too late. a new look at the identification of local average treatment effects with or without defiers. working paper, University of Southern Denmark, Dept. of Economics.

de Chaisemartin, C., 2012. All you need is late. mimeo, Paris School of Economics.

de Chaisemartin, C., 2016. Tolerating defiance? identification of treatment effects without monotonicity. working paper, University of Warwick.

de Luna, X., Johansson, P., 2014. Testing for the unconfoundedness assumption using an instrumental assumption. Journal of Causal Inference 2, 187–199.

Deaton, A. S., 2010. Instruments, randomization, and learning about development. Journal of Economic Literature 48, 424–455.

DiNardo, J., Lee, D. S., 2011. Program evaluation and research designs. In: Handbook of Labor Economics. Vol. 4. Elsevier, pp. 463–536.

Donald, S. G., Hsu, Y.-C., Lieli, R. P., 2014a. Inverse probability weighted estimation of local

average treatment effects: A higher order mse expansion. Statistics and Probability Letters 95, 132–138.

Donald, S. G., Hsu, Y.-C., Lieli, R. P., 2014b. Testing the unconfoundedness assumption via inverse probability weighted estimators of (L)ATT. Journal of Business & Economic Statistics 32 (3), 395–415.

Dzemski, A., Sarnetzki, F., 2014. Overidentification test in a nonparametric treatment model with unobserved heterogeneity. mimeo, University of Mannheim.

Fiorini, M., Stevens, K., 2014. Monotonicity in iv and fuzzy rd designs - a guide to practice. mimeo, University of Sydney.

Flores, C. A., Flores-Lagunes, A., 2013. Partial identification of local average treatment effects with an invalid instrument. Journal of Business & Economic Statistics 31, 534–545.

Flores, C. A., Flores-Lagunes, A., Chen, X., 2015. Going beyond late: Bounding average treatment effects of job corps training. IZA Discussion Paper No. 9511.

Frangakis, C., Rubin, D., 1999. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. Biometrika 86, 365–379.

Fricke, H., Frölich, M., Huber, M., Lechner, M., 2015. Endogeneity and non-response bias in treatment evaluation: Nonparametric identification of causal effects by instruments. IZA Discussion Paper No 9428.

Frölich, M., 2007. Nonparametric iv estimation of local average treatment effects with covariates. Journal of Econometrics 139, 35–75.

Frölich, M., Huber, M., 2014a. Direct and indirect treatment effects - causal chains and mediation analysis with instrumental variables. forthcoming in the Journal of the Royal Statistical Society Series B.

Frölich, M., Huber, M., 2014b. Treatment evaluation with multiple outcome periods under endogeneity and attrition. Journal of the American Statistical Association 109, 1697–1711.

Frölich, M., Lechner, M., 2015. Combining matching and nonparametric instrumental variable estimation: Theory and an application to the evaluation of active labour market policies. Journal of Applied Econometrics 30 (5), 718–738.

Frölich, M., Melly, B., 2013a. Identification of treatment effects on the treated with one-sided non-compliance. Econometric Reviews 32, 384–414.

Frölich, M., Melly, B., 2013b. Unconditional quantile treatment effects under endogeneity. Journal of Business & Economic Statistics 31, 346–357.

Frumento, P., Mealli, F., Pacini, B., Rubin, D. B., 2012. Evaluating the effect of training on wages in the presence of noncompliance, nonemployment, and missing outcome data. Journal of the American Statistical Association 107, 450–466.

Hausman, J. A., 1978. Specification tests in econometrics. Econometrica 46, 1251–71.

Heckman, J. J., 1997. Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations. The Journal of Human Resources 32, 441–462.

Heckman, J. J., Pinto, R., 2015. Unordered monotonicity. University of Chicago, mimeo.

Heckman, J. J., Schmierer, D., Urzua, S., 2010. Testing the correlated random coefficient model. Journal of Econometrics 158, 177–203.

Heckman, J. J., Urzúa, S., 2010. Comparing iv with structural models: What simple iv can and cannot identify. Journal of Econometrics 156, 27–37.

Heckman, J. J., Vytlacil, E., 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects. Proceedings National Academic Sciences USA, Economic Sciences 96, 4730–4734.

Heckman, J. J., Vytlacil, E., 2001a. Instrumental variables, selection models, and tight bounds on the average treatment effects. In: Lechner, M., Pfeiffer, M. (Eds.), Econometric Evaluation of Labour Market Policies. Center for European Economic Research, New York, pp. 1–15.

Heckman, J. J., Vytlacil, E., 2001b. Local instrumental variables. In: Hsiao, C., Morimune, K., Powell, J. (Eds.), Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya. Cambridge University Press, Cambridge.

Heckman, J. J., Vytlacil, E., May 2005. Structural equations, treatment effects, and econometric policy evaluation 1. Econometrica 73, 669–738.

Heckman, J. J., Vytlacil, E. J., 2007. Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. Vol. 6, Part B of Handbook of Econometrics. Elsevier, pp. 4875–5143.

Hernan, M. A., Robins, J. M., 2006. Instruments for causal inference. an epidemiologist's dream? Epidemiology 17, 360–372.

Hirano, K., Imbens, G. W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189.

Hong, H., Nekipelov, D., 2010. Semiparametric efficiency in nonlinear late models. Quantitative Economics 1, 279–304.

Hsu, Y.-C., Lai, T.-C., Lieli, R. P., 2015. Estimation and inference for distribution functions and quantile functions in endogenous treatment effect models, work in progress.

Huber, M., 2013. A simple test for the ignorability of non-compliance in experiments. Economics Letters 120, 389–391.

Huber, M., 2014. Sensitivity checks for the local average treatment effect. Economics Letters 123, 220–223.

Huber, M., Laffers, L., Mellace, G., 2014. Sharp iv bounds on average treatment effects on the treated and other populations under endogeneity and noncompliance. forthcoming in the Journal of Applied Econometrics.

Huber, M., Mellace, G., 2015. Testing instrument validity for late identification based on inequality moment constraints. Review of Economics and Statistics 97, 398–411.

Hull, P., 2015. Isolateing: Identifying counterfactual-specific treatment effects with cross-stratum comparisons. working paper, MIT Department of Economics.

Imbens, G. W., 2010a. Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). Journal of Economic Literature 48 (2), pp. 399–423.

Imbens, G. W., 2010b. Better late than nothing: Some comments on deaton (2009) and heckman and urzua (2009). Journal of Economic Literature 48, 399–423.

Imbens, G. W., 2014. Instrumental variables: An econometrician's perspective. IZA Discussion Paper No. 8048.

Imbens, G. W., Angrist, J., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467–475.

Imbens, G. W., Rubin, D., 1997. Estimating outcome distributions for compliers in instrumental variables models. Review of Economic Studies 64, 555–574.

Jones, D., 2015. The economics of exclusion restrictions in iv models. NBER working paper 21391, Cambridge, MA.

Kitagawa, T., 2009. Identification region of the potential outcome distribution under instrument independence. CeMMAP working paper 30/09.

Kitagawa, T., 2015. A test for instrument validity. Econometrica 83, 2043–2063.

Klein, T. J., 2010. Heterogeneous treatment effects: Instrumental variables without monotonicity? Journal of Econometrics 155, 99–116.

Kolesar, M., 2013. Estimation in an instrumental variable model with treatment effect heterogeneity. Unpublished Manuscript.

Kowalski, A. E., 2016. Doing more when you're running late: Applying marginal treatment effect methods to examine treatment effect heterogeneity in experiments. working paper, Yale University.

Lee, S., Salanie, B., 2015. Identifying effects of multivalued treatments. cemmap working paper CWP72/15.

Little, R., Rubin, D., 1987. Statistical Analysis with Missing Data. Wiley, New York.

Machado, C., Shaikh, A., Vytlacil, E., 2013. Instrumental variables, and the sign of the average treatment effect. unpublished manuscript, University of Chicago.

Maestas, N., Mullen, K. J., Strand, A., 2013. Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. The American Economic Review 103, 1797–1829.

Manski, C. F., 1990. Nonparametric bounds on treatment effects. American Economic Review, Papers and Proceedings 80, 319–323.

Mealli, F., Imbens, G., Ferro, S., Biggeri, A., 2004. Analyzing a randomized trial on breast self-examination with noncompliance and missing outcomes. Biostatistics 5, 207–222.

Mealli, F., Pacini, B., 2013. Using secondary outcomes and covariates to sharpen inference in instrumental variable settings. Journal of the American Statistical Association 108, 1120–1131.

Melly, B., Wüthrich, K., 2016. Local quantile treatment effects. prepared for the Handbook of Quantile Regression. Discussion Paper 1605, University of Bern, Department of Economics.

Miquel, R., 2002. Identification of dynamic treatment effects by instrumental variables. University of St. Gallen Economics Discussion Paper Series 2002-11.

Mogstad, M., Santos, A., Torgovitsky, A., 2016. Using instrumental variables for inference about policy relevant treatment parameters. Unpublished Manuscript.

Mourifié, I., Wan, Y., 2014. Testing late assumptions. mimeo, University of Toronto.

Richardson, T. S., Robins, J. M., 2010. Analysis of the binary instrumental variable model. In: Dechter, R., Geffner, H., Halpern, J. Y. (Eds.), Heuristics, probability and causality: a tribute to Judea Pearl. College Publications, London, UK, pp. 415–440.

Rosenbaum, P., Rubin, D., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Roy, A., 1951. Some thoughts on the distribution of earnings. Oxford Economic Papers 3, 135–146.

Rubin, D. B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66, 688–701.

Rubin, D. B., 1976. Inference and missing data. Biometrika 63, 581–592.

Shaikh, A., Vytlacil, E., 2011. Partial identification in triangular systems of equations with binary dependent variables. Econometrica 79, 949–955.

Sharma, A., 2016. Necessary and probably sufficient test for finding valid instrumental variables. working paper, Microsoft Research, New York.

Slichter, D., 2014. Testing instrument validity and identification with invalid instruments. mimeo, University of Rochester.

Small, D. S., Tan, Z., 2007. A stochastic monotonicity assumption for the instrumental variables method. Technical report, Department of Statistics, Wharton School, University of Pennsylvania.

Tan, Z., 2006. Regression and weighting methods for causal inference using instrumental variables. Journal of the American Statistical Association 101, 1607–1618.

Ura, T., 2016. Heterogeneous treatment effects with mismeasured endogenous treatment. working paper, Duke University.

Uysal, S. D., 2011. Doubly robust iv estimation of the local average treatment effects. mimeo, University of Konstanz.

Vuong, Q., Xu, H., 2014. Counterfactual mapping and individual treatment effects in nonseparable models with discrete endogeneity. Unpublished Manuscript.

Vytlacil, E., 2002. Independence, monotonicity, and latent index models: An equivalence result.

Econometrica 70, 331–341.

Wüthrich, K., 2016. A comparison of two quantile models with endogeneity. Unpublished Manuscript.

Yamamoto, T., 2013. Identification and estimation of causal mediation effects with treatment noncompliance. unpublished manuscript, MIT Department of Political Science.

Yu, P., 2014. Marginal quantile treatment effect and counterfactual analysis. Unpublished Manuscript.

Zhang, J., Rubin, D., Mealli, F., 2009. Likelihood-based analysis of causal effects of job-training programs using principal stratification. Journal of the American Statistical Association 104, 166–176.

Working Paper 479 february 2017

Authors

Martin HUBER
University of Fribourg, Department of Economics, Bd. de Pérolles 90, 1700 Fribourg, Switzerland.
Phone: +41 26 300 8274; Email: martin.huber@unifr.ch; Website: http://www.unifr.ch/appecon/en/team/martin-huber

Kaspar WÜTHRICH
UC San Diego, Department of Economics, San Diego, 9500 Gilman Dr. La Jolla, CA 92093, USA.
Phone: +1 858 534-3383 Email: kwuthrich@ucsd.edu; Website: https://sites.google.com/site/wuethricheconomics/

Abstract

This paper provides a review of methodological advancements in the evaluation of heterogeneous treatment effect models based on instrumental variable (IV) methods. We focus on models that achieve identification through a monotonicity assumption on the selection equation and analyze local average and quantile treatment effects for the subpopulation of compliers. We start with a comprehensive discussion of the binary treatment and binary instrument case which is relevant for instance in randomized experiments with imperfect compliance. We then review extensions to identification and estimation with covariates, multi-valued and multiple treatments and instruments, outcome attrition and measurement error, and the identification of direct and indirect treatment effects, among others. We also discuss testable implications and possible relaxations of the IV assumptions, approaches to extrapolate from local to global treatment effects, and the relationship to other IV approaches.

Citation proposal

Martin Huber, Kaspar Wüthrich. 2017 «Evaluating local average and quantile treatment effects under endogeneity based on instruments: a review». Working Papers SES 479, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland)

Jel Classification

C26

Keywords

Instrument, LATE, treatment effects, selection on unobservables

Working Papers SES collection

Last published

473 Deuchert E., Huber M., Schelker M.: Direct and indirect effects based on difference-in-differences with an application to political preferences following the Vietnam draft lottery; 2016
474 Grossmann V., Schäfer A., Steger T., Fuchs B.: Reversal of Migration Flows: A Fresh Look at the German Reunification; 2016
475 Pesenti A.: The Meaning of Monetary Stability; 2016
476 Furrer O., Sudharshan D., Tsiotsou Rodoula H., Liu Ben S. : A Framework for Innovative Service Design; 2016
477 Herz H., Taubinsky D.: What Makes a Price Fair? An Experimental Study of Transaction Experience and Endogenous Fairness Views; 2016
478 Zehnder C., Herz H., Bonardi J.-P.: A Productive Clash of Cultures: Injecting Economics into Leadership Research; 2016

Catalogue and download links

http://www.unifr.ch/ses/wp
http://doc.rero.ch/collection/WORKING_PAPERS_SES

Publisher

Université de Fribourg, Suisse, Faculté des sciences économiques et sociales
Universität Freiburg, Schweiz, Wirtschafts- und sozialwissenschaftliche Fakultät
University of Fribourg, Switzerland, Faculty of Economics and Social Sciences

Bd de Pérolles 90, CH-1700 Fribourg
Tél.: +41 (0) 26 300 82 00
decanat-ses@unifr.ch www.unifr.ch/ses